STEIN COUPLINGS FOR NORMAL APPROXIMATION

LOUIS H. Y. CHEN AND ADRIAN RÖLLIN

ABSTRACT. In this article we propose a general framework for normal approximation using Stein's method. We introduce the new concept of *Stein couplings* and we show that it lies at the heart of popular approaches such as the local approach, exchangeable pairs, size biasing and many other approaches. We prove several theorems with which normal approximation for the Wasserstein and Kolmogorov metrics becomes routine once a Stein coupling is found. To illustrate the versatility of our framework we give applications in Hoeffding's combinatorial central limit theorem, functionals in the classic occupancy scheme, neighbourhood statistics of point patterns with fixed number of points and functionals of the components of randomly chosen vertices of subcritical Erdős-Rényi random graphs. In all these cases, we use new, non-standard couplings.

CONTENTS

1. Introduction

Since its introduction in the early 70s, Stein's method has gone through a vivid development. First used by Stein (1972) for normal approximation of m-dependent sequences, it was gradually modified and generalized to other distributions and different dependency settings. Among the few important contributions that influenced the theoretical understanding of the method is the book by Stein (1986), where the concepts of *auxiliary randomization* and *exchangeable pairs* are introduced. Another corner stone is the *generator method*, independently introduced by Barbour (1988) and Götze (1991), which allows for an adaptation of the method to more complicated approximating distributions, such as compound Poisson distribution, Poisson point processes and Gaussian diffusions. Diaconis and Zabell (1991) give a connection between Stein's method and orthogonal polynomials; see also Goldstein and Reinert (2005a), who related this approach to distributional transformations. A more recent development was initiated by Nourdin and Peccati (2009) and related articles, where a fruitful theory for normal approximation for functionals of Gaussian measures, Rademacher sequences and Poisson measures is developed. Despite these achievements, relatively little effort has been put into building up a rigorous theoretical framework for the method in order to unify and generalize the variety of known results.

In particular for normal approximation, a wide range of different approaches has appeared over the last decades. Among the most prominent approaches are the *local approach*, dating back to Stein (1972) and extensively studied by Chen and Shao (2004), the *exchangeable pairs approach* by Stein (1986), further developed by Rinott and Rotar' (1997), Röllin (2007b) and Röllin (2008a) with a variety of applications such as weighted U-statistics, anti-voter model on finite graphs and models in statistical mechanics, and the *size* and *zero bias couplings* by Goldstein and Rinott (1996) and Goldstein and Reinert (1997), respectively. Another approach was introduced by Chatterjee (2008) for functionals of independent random variables (we will discuss a more general version of this approach—called *interpolation to independence*— in Section 3.4). In addition to these abstract approaches, many other ad-hoc constructions have been used to tackle specific problems, such as Ho and Chen (1978) and Bolthausen (1984) for the combinatorial limit theorem, Barbour, Karoński, and Ruciński (1989) and Röllin (2008b) for refined versions of local dependence and Barbour and Eagleson (1986) and Zhao, Bai, Chao, and Liang (1997) for double indexed permutation statistics. However, despite all these achievements, a unifying framework is still missing and connections between the different approaches given in the literature are vague, at best. Although making an attempt to systematically discuss Stein's method, survey articles such as Reinert (1998) and Rinott and Rotar' (2000) illustrate the key issue here: for each approach a separate theorem is proved, and then typically only for one specific metric.

The reason for this seems to be the following. For all of these approaches, the involved random variables either have to satisfy some more or less abstract conditions or some specific properties in the dependence structure are exploited to obtain the results. This could be a defining equation like in the

size-biasing approach, a linearity assumption on a conditional expectation for exchangeable pairs, a local dependence structure, or other properties and conditions. Depending on the specific form of these conditions, the quantities arising from using Stein's method—seemingly—have to be handeled differently. Although in simpler applications it might be clear how to directly manipulate these expressions in an ad-hoc way in order to successfully apply Stein's method, this is less feasible for more complex situations. Chatterjee (2008) proposes an approach for functionals of finite collections of independent random variables. His approach comes at the cost of a rather complicated bound, and for many applications optimal results (in terms of moment conditions and metrics) may not be obtained that way (c.f. the difference between Corollaries 2.2 and 2.3 below). It is crucial to exploit properties of the random variables at hand in order to express the error bounds in terms of simple and manageable expressions. To achieve this, we will explore what the abstract key conditions are that allow for a successful implementation of Stein's method for normal approximation—a question that has not yet been addressed in the literature. Our framework provides such a set of conditions along with a variety of "plug-in" type theorems. Not only does our framework show the connection between all the above mentioned approaches, but it also introduces some crucial generalisations and hence flexibility into Stein's method for normal approximation.

Our main tool is that of couplings; more specifically, we introduce the new concept of *Stein couplings*. We provide general approximation theorems with respect to the Wasserstein and Kolmogorov metrics, where the error terms are expressed in terms of the relationship between the involved coupled random variables. Although we relate the different known approaches via Stein couplings, our applications also show that the distinction usually made between these approaches is rather artificial. Most of the couplings used in our applications cannot be clearly assigned to one of the known approaches, but emerge naturally from the problem at hand and are therefore constructed in a more ad-hoc way. Nevertheless, in Section 3 we make an attempt to give a systematic overview over the different coupling constructions, being very well aware of the fact that other ways of classifying these couplings may be equally reasonable.

1.1. **A short introduction to Stein's method.** We assume throughout this article that W is a random variable whose distribution is to be approximated by a standard normal distribution and we also assume that W has finite variance. Stein's method for normal approximation is based on the fact that, for all, say, Lipschitz continuous function f we have

$$\mathbb{E}\{Zf(Z)\} = \mathbb{E}f'(Z) \qquad (1.1)$$ {1}

if and only if $Z \sim \mathrm{N}(0,1)$. Now, if it is the case that

$$\mathbb{E}\{Wf(W)\} \approx \mathbb{E}f'(W) \qquad (1.2)$$ {2}

for many functions f, we would expect that W is close to the normal distribution. With Stein's method we can make this heuristic idea rigorous in the following sense. Assume that, in order to measure the closeness of $\mathscr{L}(W)$

and $\mathscr{L}(Z)$, we would like to bound

$$\mathbb{E}h(W) - \mathbb{E}h(Z) \qquad (1.3) \quad \{3\}$$

for some function h (take for example the half line indicators for the Kolmogorov metric). Solving the so-called *Stein equation*

$$f'(x) - xf(x) = h(x) - \mathbb{E}h(Z) \qquad (1.4) \quad \{4\}$$

for $f = f_h$, we can then express the quantity (1.3) as

$$\mathbb{E}h(W) - \mathbb{E}h(Z) = \mathbb{E}\big\{f'(W) - Wf(W)\big\} = \mathbb{E}Af(W), \qquad (1.5) \quad \{5\}$$

where A is the operator defined by $Af(x) := f'(x) - xf(x)$. Hence, $\mathbb{E}Af(W)$ measures the error in (1.2) and the function f relates this error to the error of the approximation $\mathbb{E}h(W) \approx \mathbb{E}h(Z)$ via (1.5) (see Röllin (2007a) for a more detailed discussion).

Let us elaborate (1.2) more rigorously. One way to express (1.2) is to assume that there are two random variables T_1 and T_2 such that

$$\mathbb{E}\big\{Wf(W)\big\} = \mathbb{E}\big\{T_1 f'(W + T_2)\big\} \qquad (1.6) \quad \{6\}$$

for all f. Equations of the form (1.6) are often called *Stein identities*, as they characterise in some sense the distribution of W. There are two important special cases of (1.6). If, for example, Ψ is a Gaussian field and $W = W(\Psi)$ a (smooth) functional of it, Nourdin and Peccati (2009) use Maliavin calculus to derive (1.6) with $T_2 = 0$ and they give a more or less explicit expression for T_1. In contrast, in the zero-biasing approach as introduced by Goldstein and Reinert (1997), it is assumed that (1.6) holds for a specific T_2 where $T_1 = 1$.

To illustrate the line of argument to obtain a final bound in its simplest form, let us look at the case $T_2 = 0$. We can write

$$\mathbb{E}Af(W) = \mathbb{E}\big\{(1 - T_1)f'(W)\big\} = \mathbb{E}\big\{(1 - \mathbb{E}^W T_1)f'(W)\big\}, \qquad (1.7) \quad \{7\}$$

so that the error in the normal approximation is given by

$$|\mathbb{E}h(W) - \mathbb{E}h(Z)| = |\mathbb{E}Af(W)| \leqslant \|f'\|\mathbb{E}|1 - \mathbb{E}^W T_1|,$$

and, if $\mathbb{E}T_1 = 1$, the last term is usually further bounded by $\sqrt{\operatorname{Var}\mathbb{E}^W T_1}$. It is not difficult to show that $\|f'\| \leqslant 2\|h\|$ (where $\|\cdot\|$ denotes the supremum norm). Hence, we obtain the final bound

$$|\mathbb{E}h(W) - \mathbb{E}h(Z)| \leqslant 2\|h\|\sqrt{\operatorname{Var}\mathbb{E}^W T_1}.$$

Note that this specific form of the bound involving $\operatorname{Var}\mathbb{E}^W T_1$ has been implicitly used in the literature around Stein's method many times, but probably first made explicit by Cacoullos, Papathanasiou, and Utev (1994). In Chatterjee (2009) and Nourdin, Peccati, and Reinert (2009) a connection with Poincaré inequalities was made, where a bound on $\operatorname{Var}\mathbb{E}^W T_1$ is called a *Poincaré inequality of second order*.

1.2. **An outline of our approach.** Roughly speaking, we propose a general, probabilistic method of deriving identities of the form (1.6) and more refined versions of it, where we will make use of both random variables T_1 and T_2 as we will show below. Based on this, we provide theorems to obtain bounds for the accuracy of a normal approximation of W.

The key idea is that of *auxiliary randomisation*, introduced by Stein (1986). To this end we construct a random variable W' which we imagine to be a 'small perturbation' of W. It is important to emphasize that we make no assumptions about the distribution of W' or the joint distribution at this point (in particular no exchangeability is assumed a priori). We also need to note that we never attempt to couple W and W' so that $W' = W$ almost surely; in fact, such a coupling would contain no useful information for our purposes. It is crucial that *some* randomness remains (see Ross (to appear) for a discussion on how to optimally choose the perturbation in some examples using exchangeable pairs) and it will become clear that the difference $D := W' - W$ contains essential information about W. This can be seen as a *local-to-global* approach, where we deduce global properties of W from the behaviour of local perturbations, as these are often easier to handle.

When dealing with Stein's method, it becomes clear that we cannot expect normal approximation results from any arbitrary coupling (W, W'), and we need to impose some structure. To achieve this, we generalize an idea which goes back to Stein (1972), and introduce a third random variable G. For reasons that will hopefully become apparent in the course of this article, we then make the following key definition.

Definition 1.1. Let (W, W', G) be a coupling of square integrable random variables. We call (W, W', G) a *Stein coupling* if

$$\mathbb{E}\{Gf(W') - Gf(W)\} = \mathbb{E}\{Wf(W)\} \qquad (1.8) \quad \{8\}$$

for all functions for which the expectations exist.

Before explaining how this will help in finding an identity of the form (1.6), let us first discuss some standard Stein couplings. As a simple example where (1.8) holds, assume that (W, W') is an exchangeable pair and assume that, for some $\lambda > 0$ we have

$$\mathbb{E}^W(W' - W) = -\lambda W. \qquad (1.9) \quad \{9\}$$

If we set $G = \frac{1}{2\lambda}(W' - W)$ it is easy to see that (1.8) is satisfied (see Section 3.1 for more details). Equation (1.9) is the well-known linear regression condition introduced by Stein in Diaconis (1977) and Stein (1986) and (1.8) can be seen as a generalization of it. We will show in Section 3 that (1.8) is the key to normal approximation using Stein's method and that many approaches in the literature in fact (implicitly) establish (1.8).

Remark 1.2. Let (W, W', G) be a Stein coupling. If we choose $f(x) = 1$, we see from (1.8) that $\mathbb{E}W = 0$. If we choose $f(x) = x$ we furthermore have that $\mathbb{E}(GD) = \operatorname{Var} W$.

Note that the statement $\mathbb{E}(GD) = \operatorname{Var} W$ is well known in a special case. If W' is an independent copy of W, then $(W, W', (W' - W)/2)$ is a Stein

coupling (use the exchangeable pairs approach above with $\lambda = 1$) and, hence,

$$\operatorname{Var} W = \tfrac{1}{2}\mathbb{E}(W - W')^2 = \mathbb{E}(GD)$$

is the well known way to express the variance of W in terms of two independent copies. However, this coupling is not useful for our purpose as $|W'-W|$ is not 'small'. This example also shows that a Stein coupling by itself does by no means guarantee proximity to the normal distribution.

It is often not difficult to construct a 'small perturbation' W' of W. Three main techniques have been used in the literature: deletion, replacement and duplication (note, however, that this is often done implicitly and not expressed in terms of couplings). In many typical situations, W is a functional of a family of random variables $X_1, \ldots, X_n$ and W' can be constructed by picking a random index I independently of the X_i and then by perturbing at position I, either by removing X_I, replacing it, or by adding another, related random variable. If the X_i are not independent, the other random variables typically have to be 'adjusted' appropriately. Let us quickly illustrate the three techniques in the most simple situation, that of a sum of independent random variables. Let $W = \sum_{i=1}^{n} X_i$, where the X_i are assumed to be independent, centred and such that $\operatorname{Var} X_i = 1/n$. Let in what follows I be uniformly distributed over $\{1, \ldots, n\}$ and independent of all else.

Deletion. Define $W' = W - X_I$, that is, remove X_I from W. If we choose $G = -nX_I$, we have

$$\mathbb{E}\{Gf(W')\} = -\sum_{i=1}^{n} \mathbb{E}\{X_i f(W - X_i)\} = 0$$

due to the independence assumption. Further,

$$-\mathbb{E}\{Gf(W)\} = \sum_{i=1}^{n} \mathbb{E}\{X_i f(W)\} = \mathbb{E}\{Wf(W)\},$$

so that, indeed, (1.8) is satisfied. This construction is very powerful under local dependence, but it can also be used in other contexts; see Section 4.1.

Replacement. Let $X_1', \ldots, X_n'$ be independent copies of the X_i. Define $W' = W - X_I + X_I'$. Then, it is not difficult to see that (W, W') is an exchangeable pair and that (1.9) holds with $\lambda = 1/n$, which corresponds to $G = \frac{n}{2}(X_I' - X_I)$; this implies (1.8). The idea of replacing (or re-sampling) is one of the most fruitful in Stein's method and will often lead to an exchangeable pair (W, W') and can be applied in situations where the functional is no longer a sum or where some weak global dependence structure is present. It lends itself naturally if W can be interpreted as the state of a stationary Markov chain and W' is a step ahead in the chain; this observation was first made explicit by Rinott and Rotar' (1997). It is important to note here that the choice of G is by far not restricted to be a multiple of $(W' - W)$, also if (W, W') is exchangeable. This added flexibility is one of the key observation in this article. Note also that the X_i' need not be copies of the X_i and may as well have a different distribution, so that (W, W') need not be exchangeable. In the size-biasing approach, indeed, the X_i' will typically be chosen to have the size-biased distribution of X_i.

Duplication. To the best of our knowledge, this method has been only used by Chen (1998). Let the X_i' be as in the previous paragraph, and let $W' = W + X_I'$ along with $G = n(X_I' - X_I)$. Due to symmetry we have

$$\mathbb{E}\big\{X_I'f(W')\big\} = \mathbb{E}\big\{X_I f(W')\big\},$$

hence $\mathbb{E}\big\{Gf(W')\big\} = 0$. Further, $\mathbb{E}\big\{X_I'f(W)\big\} = 0$, so that $\mathbb{E}\big\{Gf(W)\big\} = \mathbb{E}\big\{Wf(W)\big\}$ and (1.8) follows.

As can be readily seen from this example, there are typically many possible ways to construct Stein couplings and it depends on the application which perturbation will give optimal results. Typically, two of the three random variables W, W' and G are easy to construct (usually (W, W') or (W, G)) and the challenge lies then in constructing the third random variable to make the triple a Stein coupling. However, as we will show in Section 3, by making abstract assumptions about the structure of W, many situations can be handled by standard couplings, so that, in a concrete application, one often only needs to concentrate on constructing the coupling satisfying some abstract conditions rather than to find a Stein coupling from scratch—although the latter can give interesting additional insight into the problem at hand and may also lead to improved bounds.

We need to emphasize at this point that our abstract theorems will also hold if (1.8) is not satisfied, so that—a priori—any coupling (W, W', G) can be used. However, useful bounds can only be expected if (1.8) holds at least approximately and the accuracy at which (1.8) holds enters explicitly into our error bounds. This parallels the introduction of a remainder term in Condition (1.9) by Rinott and Rotar' (1997, Eq. (1.7)).

Let us now go back and show how a coupling (W, W', G) helps in obtaining a Stein identity of the form (1.6). By the fundamental theorem of calculus we have

$$f(W') - f(W) = \int_0^D f'(W + t)dt, \qquad (1.10) \quad \{11\}$$

so that, multiplying (1.10) by G and taking expectation, we obtain

$$\mathbb{E}\big\{Gf(W') - Gf(W)\big\} = \mathbb{E}\Big\{G\int_0^D f'(W + t)dt\Big\}. \qquad (1.11) \quad \{12\}$$

If U is an independent random variable with uniform distribution on $[0, 1]$ we can also write this as

$$\mathbb{E}\big\{Gf(W') - Gf(W)\big\} = \mathbb{E}\big\{GDf'(W + UD)\big\}. \qquad (1.12) \quad \{13\}$$

If (W, W', G) is a Stein coupling, the left hand side of (1.12) equals to $\mathbb{E}\big\{Wf(W)\big\}$ and, hence, (1.6) is satisfied with $T_1 = GD$ and $T_2 = UD$. Our generality comes at a cost: as is clear from (1.6), if T_2 is non-trivial we can not easily condition T_1 on W (or an appropriate larger σ-algebra) as done in (1.7)—an important step in the argument. However, using a simple Taylor expansion, we will see how to circumvent this problem. We remark that it is usually not necessary to condition T_1 exactly on W—it is typically enough to condition on a larger σ-algebra. However, some form of conditioning is usually necessary (typically averaging over all possible 'small parts' that can be perturbed).

We note at this point that there is an infinitesimal version of the perturbation idea, introduced by Stein (1995) and further elaborated by Meckes (2008). It is applicable if the underlying random variables are continuous. Starting with an identity of the form (1.6) with non-trivial $T_1(\varepsilon)$ and $T_2(\varepsilon)$ (in the original work by means of classic exchangeable pairs) depending on some $\varepsilon > 0$, a preceding limit argument $\varepsilon \to 0$ yields an identity (1.6) with non-trivial T_1 but $T_2 = 0$.

Having a coupling (W, W', G) at hand one is already in a position to apply our theorems and obtain bounds of closeness to normality in the respective metric. However, bounding $\operatorname{Var} \mathbb{E}^W T_1 = \operatorname{Var} \mathbb{E}^W (GD)$ is not always easy and sometimes not optimal as one typically has to make use of (truncated) fourth moments of the involved random variables. For this reason it is often beneficial—and sometimes crucial—to introduce other auxiliary random variables.

The first extension is to replace conditioning on W by conditioning on another random variable W'', which is still assumed to be close to W but typically independent of GD (but not independent of W and W'!). In this case the main error term becomes $\operatorname{Var}^{W''}(GD)$ which of course vanishes if W'' is independent of GD. Although this comes at the cost of additional error terms, these are usually easier to bound. We need to emphasize that the distribution of W'' is—a priori—irrelevant and no corresponding equation of the form (1.8) has to be satisfied for W''; the only important feature is independence from GD and closeness to W. In fact, we can show that, if (W, W', G) is a Stein coupling and W'' is independent of GD, then we can construct T_3 and T_4 such that

$$\mathbb{E}\{W f(W)\} = \mathbb{E} f'(W) + \mathbb{E}\{T_3 f''(W + T_4)\}$$

for all smooth enough functions f. Compared to (1.6), this is clearly a step further towards (1.1). Typically, the specific dependence structure in W used to construct W'' independently of GD can also be exploited to calculate bounds on $\operatorname{Var} \mathbb{E}^W (GD)$—so why introducing W'' in the first place? The crucial advantage is that the dependence structure can be exploited in a more direct way (that is, in an earlier stage of the proof), avoiding forth moments—and constructing W'' is often easier than bounding $\operatorname{Var} \mathbb{E}^W (GD)$.

Other improvements can be made in specific applications by replacing D by a random variable $\tilde{D}$ such that $\mathbb{E}^W (G\tilde{D}) = \mathbb{E}^W (GD)$ (note that we use the same letter only for convenience; $\tilde{D}$ itself does not have to be the difference of two random variables) and replacing 1 in (1.7) by a random variable S such that $\mathbb{E}^W S = 1$. Smart choices of $\tilde{D}$ and S may allow us to construct W'' to be closer to W in order to improve or simplify the error bounds. As we will see in Section 3.2.2 about decomposable random variables, the use of $\tilde{D}$ can be crucial. However, at first reading one may always assume that $W'' = W$, $\tilde{D} = D$ and $S = 1$.

We do not claim that *all* results that have been obtained using Stein's method for normal approximation can be represented in terms of these auxiliary random variables; but we provide evidence that we can cover a large part of them. It may be possible to setup an even more general framework by

introducing other auxiliary random variables so that even very specialised results such as Sunklodas (2008), where very fine and delicate calculations are necessary to obtain optimal rates, could be represented in such a framework. We did not attempt to do so in order to keep our theorems manageable.

As mentioned after (1.6), zero-biasing couplings, as introduced in Goldstein and Reinert (1997), are the special case of (1.6) where $T_1 = 1$. It is thus not surprising, that we are not able to directly represent a zero-bias coupling as a Stein coupling. However, we will show that each coupling (W, W', G) satisfying (1.8) gives rise to a zero-bias construction (but, unfortunately, the construction does not directly lead to a coupling with W). Based on an exchangeable pair, Goldstein and Reinert (2005b) propose a way to construct the zero bias W^z. We adapt this construction to our more general setting. As such, the zero-bias approach *parallels* our approach rather than being a special case of it. Therefore, in this article, we take a different point of view than, for example, Goldstein and Reinert (1997) and Goldstein and Reinert (2005a), where size and zero biasing are seen to be closely related (both are distributional transformations). From our perspective, size biasing is closer related to approaches such as local approach and exchangeable pairs approach and we think of zero biasing as being separate from these.

The rest of the article is organized as follows. In the remainder of the introduction we will introduce the metrics of interest. In Section 2 we will present the main theorems of the article and discuss the crucial error terms. In Section 3 we will show how known approaches fit into our framework and also present and discuss new couplings. Section 4 is dedicated to some applications in order to see different couplings in action. In Section 5 we will make the connection with the zero-bias approach and in Section 6 we will prove the main results from Section 2.

1.3. **The probability metrics.** For probability distribution functions P and Q define

$$d_{\mathrm{W}}(P, Q) = \int_{-\infty}^{\infty} |P(x) - Q(x)| dx, \qquad d_{\mathrm{K}}(P, Q) = \sup_{x \in \mathbb{R}} |P(x) - Q(x)|.$$

The first quantity is known as L_1, *Wasserstein* or *Kantorovich* metric and is only a metric on the set of probability distributions with finite first moment. If $X \sim P$ and $Y \sim Q$ have finite first moments and if $\mathcal{F}_{\mathrm{W}}$ is the set of Lipschitz continuous functions on $\mathbb{R}$ with Lipschitz constant at most 1, we have

$$d_{\mathrm{W}}(P, Q) = \sup_{h \in \mathcal{F}_{\mathrm{W}}} \big| \mathbb{E}h(X) - \mathbb{E}h(Y) \big|,$$

where the infimum ranges over all possible couplings of X and Y. The second metric is known as *Kolmogorov* or *uniform metric* and if $\mathcal{F}_{\mathrm{K}}$ denotes the set of half line indicators we obviously have

$$d_{\mathrm{K}}(P, Q) = \sup_{h \in \mathcal{F}_{\mathrm{K}}} \big| \mathbb{E}h(X) - \mathbb{E}h(Y) \big|.$$

If φ is a right continuous function on $\mathbb{R}$ such that for each ε we have

$$Q(B^{\varepsilon}) \leqslant Q(B) + \varphi(\varepsilon)$$

for all Borel set $B \subset \mathbb{R}$, where $B^\varepsilon = \{x : \inf_{y \in B} |x - y| \leqslant \varepsilon\}$, then

$$d_{\mathrm{K}}(P, Q) \leqslant \sqrt{d_{\mathrm{W}}(P, Q) + \varphi(d_{\mathrm{W}}(P, Q))}$$

(see Gibbs and Su (2002) for a compilation of this and similar results). If $Q = \mathrm{N}(0, 1)$ is the standard normal measure we can take $\varphi(x) = x\sqrt{2/\pi}$, thus

$$d_{\mathrm{K}}(P, \mathrm{N}(0, 1)) \leqslant 1.35\sqrt{d_{\mathrm{W}}(P, \mathrm{N}(0, 1))}. \qquad (1.13) \quad \{14\}$$

However, as is well known, this bound is often not optimal and, in fact, in many situations both metrics will exhibit the same rates of convergence.

2. Main results

Throughout this section, let all random variables be at least square integrable and defined on the same probability space without making any further assumptions unless explicitly stated. Again, we point out that at first reading one may set $W'' = W$, $\tilde{D} = D$ and $S = 1$. It may also be helpful to keep in mind the introductory couplings for sums of independent random variables in Section 1.2. Before stating our main theorems, we first define and discuss some error terms in order to express the overall error bounds in the different metrics and using different techniques. Let

$$r_0 = \sup_{\|f\|, \|f'\| \leqslant 1} \big| \mathbb{E}\{Gf(W') - Gf(W) - Wf(W)\} \big|, \qquad (2.1) \quad \{15\}$$

where the supremum is meant to be taken over all function f which are bounded by 1 and Lipschitz continuous with Lipschitz constant at most 1. We need to point out that in the proofs the actual supremum is only taken over the solutions to the Stein equation so that in cases where more properties of f are needed (such as better constants) this can easily be accomplished. Clearly, if (W, W', G) is a Stein coupling, then $r_0 = 0$. Hence, for the couplings and examples discussed in this article the actual set of functions over which the supremum is taken is not relevant. In cases where the linearity condition is not exactly satisfied (see e.g. Rinott and Rotar' (1997) and Shao (2005)) r_0 measures the corresponding error; Shao (2005) handles self-normalised sums where uniformly bounded derivatives of f are needed to proof that (the implicitly used) r_0 is small. Let now

$$D := W' - W, \qquad D' := W'' - W,$$

let $\tilde{D}$ be a square integrable and let S be an integrable random variable on the same probability space. Define

$$r_1 = \mathbb{E}\big| \mathbb{E}^W (GD - G\tilde{D}) \big|, \quad r_2 = \mathbb{E}\big| \mathbb{E}^W (1 - S) \big|, \quad r_3 = \mathbb{E}\big| \mathbb{E}^{W''} (G\tilde{D} - S) \big|.$$

Clearly, r_1 is the error we make by replacing D by $\tilde{D}$, r_2 the error we make by replacing 1 by S and r_3 corresponds to the main error term as discussed in the introduction. These error terms will appear irrespective of the metric. Additional error terms which are specific to the metric of interest will be defined in the respective sections.

2.1. **Wasserstein distance.** Bounds for smooth test functions are typically easier to obtain as no smoothness of W is required. Let us first start with a very general theorem, from which we will then deduce some simpler corollaries.

Theorem 2.1. *Let W, W', W'', G and $\tilde{D}$ be square integrable random variables and let S be an integrable random variable. Then*

$$d_{\mathrm{W}}\big(\mathscr{L}(W), \mathrm{N}(0,1)\big)$$
$$\leqslant 2r_0 + 0.8r_1 + 0.8r_2 + 0.8r_3 + 1.6r_4 + r_5 + 1.6r_4' + 2r_5',$$

where

$$r_4 = \mathbb{E}|GD\,\mathrm{I}[|D| > 1]|, \qquad r_4' = \mathbb{E}|(G\tilde{D} - S)\mathrm{I}[|D'| > 1]|,$$
$$r_5 = \mathbb{E}|G(D^2 \wedge 1)|, \qquad r_5' = \mathbb{E}|(G\tilde{D} - S)(|D'| \wedge 1)|.$$

Note that, here and in later results, the truncation constant 1 is not chosen arbitrarily as the truncation at 1 has some optimality properties; see Loh (1975) and Chen and Shao (2001). The difference $|G\tilde{D} - S|$ in r_4' and r_5' can usually be replaced by $|G\tilde{D}| + |S|$ without much loss of precision.

Under additional, but not too strong assumptions we have the following.

Corollary 2.2. *Let (W, W', G) be a Stein coupling with $\mathrm{Var}\,W = 1$. Then, under finite fourth moments assumption,*

$$d_{\mathrm{W}}\big(\mathscr{L}(W), \mathrm{N}(0,1)\big) \leqslant 0.8\sqrt{\mathrm{Var}\,\mathbb{E}^W(GD)} + \mathbb{E}|GD^2|. \qquad (2.2) \quad \{16\}$$

Proof. Theorem 2.1 with $S = 1$, $\tilde{D} = D$ and $W'' = W$ yields the result, but with bigger constants. However, with the assumption of finite fourth moments no truncation is necessary, and one can obtain the better result (2.2) with essentially the same proof as for Theorem 2.1. $\qquad\square$

Corollary 2.3. *Let (W, W', G) be a Stein coupling with $\mathrm{Var}\,W = 1$ and assume that there are S and $\tilde{D}$ such that $\mathbb{E}^W S = 1$, $\mathbb{E}^W(G\tilde{D}) = \mathbb{E}^W(GD)$ and W'' independent of $(G\tilde{D}, S)$. Then, under finite third moments of W, W', G and $\tilde{D}$ and $\mathbb{E}|S|^{3/2} < \infty$,*

$$d_{\mathrm{W}}\big(\mathscr{L}(W), \mathrm{N}(0,1)\big) \leqslant \mathbb{E}|GD^2| + 2\mathbb{E}|G\tilde{D}D'| + 2\mathbb{E}|SD'|.$$

Using Hölder's inequality, we obtain the following straightforward simplification which gives the correct order in 'typical' situations; we assume $S = 1$.

Corollary 2.4. *Let (W, W', G) be a Stein coupling with $\mathrm{Var}\,W = 1$ and assume that there is $\tilde{D}$ such that $\mathbb{E}^W(G\tilde{D}) = \mathbb{E}^W(GD)$ and W'' independent of $G\tilde{D}$. Then, if*

$$\mathbb{E}|D|^3 \vee \mathbb{E}|\tilde{D}|^3 \vee \mathbb{E}|D'|^3 \leqslant A^3, \qquad \mathbb{E}|G|^3 \leqslant B^3, \qquad (2.3) \quad \{17\}$$

for some positive constants A and B, we have

$$d_{\mathrm{W}}\big(\mathscr{L}(W), \mathrm{N}(0,1)\big) \leqslant 5A^2B. \qquad (2.4) \quad \{18\}$$

Proof. The only part which may need explanation is that $\mathbb{E}(GD) = 1$ because $\operatorname{Var} W = 1$ by assumption and hence $1 \leqslant \mathbb{E}|GD|$. This yields

$$\mathbb{E}|D'| \leqslant \mathbb{E}|GD|\mathbb{E}|D'|.$$

The remaining part is due to Hölder's inequality.   □

2.2. **Kolmogorov distance.** Let us now consider bounds with respect to the Kolmogorov metric. To obtain such results, all approaches using Stein's method will estimate probabilities of the form

$$\mathbb{P}[A \leqslant W \leqslant B \,|\, \mathcal{F}] \qquad\qquad (2.5) \qquad \{19\}$$

at some stage of the proof, where A and B are $\mathcal{F}$-measurable random variables. This is in order to control the smoothness of W as we are now dealing with a non-smooth metric. Essentially two techniques can be found in the literature to deal with such expressions. In the first approach, bounds on (2.5) are established by using similar techniques as used for Stein's method, i.e. Identity (1.11) is applied for special functions f to obtain an explicit bound on (2.5). We call this the *concentration inequality approach*; see Ho and Chen (1978), Chen and Shao (2005), Shao and Su (2006), Chatterjee, Fulman, and Röllin (2006). In the second approach, (2.5) is bounded in an indirect way in terms of $d_{\mathrm{K}}\big(\mathscr{L}(W|\mathcal{F}), \mathrm{N}(\mu_{\mathcal{F}}, \sigma_{\mathcal{F}}^2)\big)$ where $\mu_{\mathcal{F}}$ and $\sigma_{\mathcal{F}}^2$ are the conditional expectation and variance, respectively, of W (see Lemma 6.2 in Section 6) which will lead to some form of recursive inequality; hence we will call this the *recursive approach*. Such inequalities are either solved directly, i.e. if $\mathcal{F}$ is the trivial σ-algebra (see Rinott and Rotar′ (1997) and Raič (2003)), or by using an inductive argument, making use of some additional structure in W; this is typically the case if $\mathcal{F}$ is a non-trivial σ-algebra (see Bolthausen (1984)). The recursive approach has the advantage that an explicit bound of (2.5) is not needed. This comes at cost of more structure in the coupling.

As has been observed by many authors, such as Rinott and Rotar′ (1997), Raič (2003) or Chen and Shao (2005) in the context of Stein's method, it is easier to obtain Kolmogorov bounds under some boundedness conditions, in which case the recursive approach can be easily implemented, in fact, with $\mathcal{F}$ equal to the trivial σ-algebra. Using a truncation argument, boundedness can be relaxed, but in order to obtain useful results one will need fast decaying tails of G, D, $\tilde{D}$ and D'. We will use mainly this approach for our applications; see also for example Shao and Su (2006) and Chatterjee et al. (2006).

Theorem 2.5. *Let W, W', G, $\tilde{D}$, W'' be square integrable random variables and S be an integrable random variable. Then, for any non-negative constants α, β, β', $\tilde{\beta}$ and γ*

$d_{\mathrm{K}}\big(\mathscr{L}(W), \mathrm{N}(0,1)\big)$

$\quad \leqslant 2\big(r_0 + r_1 + r_2 + r_3 + r_6 + r_6' + (\alpha\tilde{\beta} + \gamma)(\mathbb{E}|W| + 5)\beta' + (\mathbb{E}|W| + 3)\alpha\beta^2\big).$

where

$\qquad r_6 = \mathbb{E}\big|GD\,\mathrm{I}[|G| > \alpha \text{ or } |D| > \beta]\big|,$

$\qquad r_6' = \mathbb{E}\big|(G\tilde{D} - S)\mathrm{I}[|G| > \alpha \text{ or } |\tilde{D}| > \tilde{\beta} \text{ or } |D'| > \beta' \text{ or } |S| > \gamma]\big|.$

If a sequence of couplings $(W_n, W'_n, G_n, W''_n, \tilde{D}_n, S_n)_{n \geqslant 1}$ is under consideration, the truncation points α, β, β', $\tilde{\beta}$ and γ will of course need to dependent on n. In a typical situation, say a sum of n bounded i.i.d. random variables, we will have $\alpha \asymp n^{1/2}$, $\beta \asymp \beta' \asymp \tilde{\beta} \asymp n^{-1/2}$ and $\gamma \asymp 1$.

Corollary 2.6. *Let (W, W', G) be a Stein coupling with $\operatorname{Var} W = 1$. If G and D are bounded by positive constants α and β, respectively, then*

$$d_{\mathrm{K}}\big(\mathscr{L}(W), \mathrm{N}(0,1)\big) \leqslant 2\sqrt{\operatorname{Var} \mathbb{E}^W(GD)} + 8\alpha\beta^2$$

Note that, even if G and D are bounded, we unfortunately cannot deduce a direct, useful bound on $\operatorname{Var} \mathbb{E}^W(GD)$ from that fact. Instead, we need again more structure in order to avoid $\operatorname{Var} \mathbb{E}^W(GD)$.

Corollary 2.7. *Let (W, W', G) be a Stein coupling with $\operatorname{Var} W = 1$ and assume that there are $\tilde{D}$ such that $\mathbb{E}^W(G\tilde{D}) = \mathbb{E}^W(GD)$, S such that $\mathbb{E}^W S = 1$ and W'' independent of $(G\tilde{D}, S)$. If the absolute values of G, D, $\tilde{D}$, D' and S are bounded by α, β, $\tilde{\beta}$, β' and γ, respectively, then*

$$d_{\mathrm{K}}\big(\mathscr{L}(W), \mathrm{N}(0,1)\big) \leqslant 8\alpha\beta^2 + 12\alpha\tilde{\beta}\beta' + 12\gamma\beta'.$$

We need to emphasize the remarkable statement of Corollary 2.7: under the conditions stated we immediately obtain a bound on the Kolmogorov distance to the standard normal without any additional computations! Examples can easily found such as bounded, locally dependent random variables; see Section 3.2.

The approach we will use for the next theorem was developed by Chen and Shao (2004) for locally dependent random variables. Although a concentration inequality approach was used by Chen and Shao (2004), the recursive approach is easy to implement without loss of precision. Like in Theorem 2.5, the aim is to obtain a bound involving (2.5) with respect to the unconditional W. This comes at the cost of truncated forth moments, especially in the form of r_8. Hence, the approach of avoiding truncated forth moments by making use of W'' in r_3 will not be useful because of the presence of r_8. Therefore, we give below only a version for $W'' = W$, $\tilde{D} = D$ and $S = 1$ to avoid unnecessary overloading of the bound. To define some additional error terms, let

$$\hat{K}(t) := G(\mathrm{I}[0 \leqslant t < D] - \mathrm{I}[D \leqslant t < 0]), \qquad (2.6) \quad \{20\}$$
$$K^W(t) := \mathbb{E}^W \hat{K}(t), \qquad K(t) := \mathbb{E}\hat{K}(t).$$

Theorem 2.8. *Let W, W' and G be square integrable random variables on the same probability space. Then*

$$d_{\mathrm{K}}\big(\mathscr{L}(W), \mathrm{N}(0,1)\big) \leqslant 2r_0 + 2\hat{r}_3 + 2r_4 + 2(\mathbb{E}|W| + 2.4)r_5$$
$$+ 1.4r_7 + 2((\mathbb{E}(|W| + 1)^2)^{1/2} + 1.1)r_8$$

where $\hat{r}_3 = \mathbb{E}\big|\mathbb{E}^W(GD) - 1\big|$, where r_4 and r_5 are defined as in Theorem 2.1 and where

$$r_7 = \int_{|t| \leqslant 1} \operatorname{Var} K^W(t)\,dt, \qquad r_8 = \left(\int_{|t| \leqslant 1} |t| \operatorname{Var} K^W(t)\,dt \right)^{1/2}.$$

Let us discuss $\operatorname{Var} K^W(t)$. Typically, W will consist of n parts, such as a sum of n random variables or a functional in n coordinates. In this case, the perturbation W' will typically be constructed by picking a small part of W, say part I, where I is uniform on $\{1, \ldots, n\}$ and then perturb this part. Thus, we typically will have $G := nY_I$, $D := D_I$ for sequences $Y_1, \ldots, Y_n$ and $D_1, \ldots, D_n$ and then define $W' := W + D_I$, where, with $\sigma^2 = \operatorname{Var} W$, $\mathbb{E}|Y_i| = \mathrm{O}(\sigma^{-1/2})$ and $\mathbb{E}|D_i| = \mathrm{O}(\sigma^{-1/2})$. Let now $(W^*, W^{*\prime}, G^*)$ be an independent copy of (W, W', G) and let $\mathrm{I}_t^{\pm}(x) = \mathrm{I}[0 \leqslant t < x] - \mathrm{I}[x \leqslant t < 0]$. Then we can write

$$\operatorname{Var} K^W(t) \leqslant \sum_{i,j=1}^{n} \operatorname{Cov}(Y_i \mathrm{I}_t^{\pm}(D_i), Y_j \mathrm{I}_t^{\pm}(D_j))$$

$$= \sum_{i,j=1}^{n} \mathbb{E}\big\{ Y_i Y_j \mathrm{I}_t^{\pm}(D_i) \mathrm{I}_t^{\pm}(D_j) - Y_i Y_j^* \mathrm{I}_t^{\pm}(D_i) \mathrm{I}_t^{\pm}(D_j^*) \big\}$$

so that

$$r_7 \leqslant \sum_{i,j=1}^{n} \mathbb{E}\big\{ Y_i Y_j \mathrm{I}[D_i D_j > 0](|D_i| \wedge |D_j| \wedge 1)$$

$$- Y_i Y_j^* \mathrm{I}[D_i D_j^* > 0](|D_i| \wedge |D_j^*| \wedge 1) \big\}$$

and

$$r_8^2 \leqslant \frac{1}{2} \sum_{i,j=1}^{n} \mathbb{E}\big\{ Y_i Y_j \mathrm{I}[D_i D_j > 0](|D_i|^2 \wedge |D_j|^2 \wedge 1)$$

$$- Y_i Y_j^* \mathrm{I}[D_i D_j^* > 0](|D_i|^2 \wedge |D_j^*|^2 \wedge 1) \big\},$$

respectively. In the case of local dependence, these quantities can now be bounded relatively easily; see Chen and Shao (2004).

If truncated fourth moments are to be avoided and no boundedness can be assumed, it seems that more structure is needed in the coupling. A typical instance is the use of higher-order neighbourhoods under local dependence as in Chen and Shao (2004), or the recursive structure in the combinatorial CLT in Bolthausen (1984). The theorem below is the basis for such results and it contains expressions of the form (2.5) explicitly, so that further steps are needed for a final bound.

Define for a random element X defined on the same probability space as W the quantity

$$\vartheta_\varepsilon(X) = \sup_{a \in \mathbb{R}} \mathbb{P}[a \leqslant W \leqslant a + \varepsilon \mid X],$$

where we assume without further mentioning that the regular conditional probability exists.

Theorem 2.9. *Let W, W', W'', $\tilde{D}$ and G be random variables with finite third moments and S be a random variable with $\mathbb{E}|S|^{3/2} < \infty$. Then, for any $\varepsilon > 0$,*

$$d_{\mathrm{K}}\big(\mathscr{L}(W), \mathrm{N}(0,1)\big)$$

$$\leqslant r_0 + r_1 + r_2 + r_3 + r_9 + 0.5 r_{10} + \varepsilon^{-1} r_{11}(\varepsilon) + 0.5 \varepsilon^{-1} r_{12}(\varepsilon) + 0.4\varepsilon$$

$$\tag{2.7}$$

{21}

where

$$r_9 = \mathbb{E}\big|(S - G\tilde{D})(|W| + 1)(|D'| \wedge 1)\big| \qquad r_{10} = \mathbb{E}\big|G(|W| + 1)(D^2 \wedge 1)\big|$$

$$r_{11}(\varepsilon) = \mathbb{E}\big|(S - G\tilde{D})D'\vartheta_\varepsilon(G, \tilde{D}, D', S)\big| \quad r_{12}(\varepsilon) = \mathbb{E}\big|GD^2\vartheta_\varepsilon(G, D)\big|$$

We now look at a method to obtain a final bound from the above theorem using induction. Although never mentioned in the literature around Stein's method, this type of argument can be traced back to Bergström (1944), who uses Lindeberg's method and an inductive argument to prove a Kolmogorov bound in the CLT. The argument was used later by Bolthausen (1982) in the context of martingale central limit theorems. The following Lemma 2.10 provides the key element to the inductive approach in the context of Stein's method as introduced by Bolthausen (1984). It can be used to obtain a final bound from an estimate of the form (2.7), provided that $W = W_n$ has some recursive structure and S_ε can be expressed in terms of the closeness of $W_1, W_2, \ldots, W_{n-1}$ to the standard normal distribution. Note that in the following lemma, the numbers κ_k, $k = 1, \ldots, n$ denote the respective bounds on the Kolmogorov distance between W_k and the standard normal. Whereas Bolthausen (1984) uses a recursion involving $\kappa_{n-4}, \ldots, \kappa_{n-1}$, Goldstein (2010) introduces a version involving all possible $\kappa_1, \ldots, \kappa_{n-1}$ to prove Berry-Esseen type bounds for degree counts in the Erdős-Rényi random graph using size biasing. Incidentally, already Bergström (1944) uses $\kappa_1, \ldots, \kappa_{n-1}$ for his inductive argument, although his argument is of a somewhat different flavour. The following lemma is inspired by the work of Goldstein (2010), but adapted to be used along with Theorem 2.9. An independent proof will be given in Section 6.

Lemma 2.10. *Let $\kappa_1, \ldots, \kappa_n$, be a sequence of non-negative numbers such that $\kappa_1 \leqslant 1$. Assume that there is a constant $A \geqslant 0$, a triangular array $A_{k,1}, \ldots, A_{k,k} \geqslant 0$, $k = 2, 3, \ldots, n$, and a sequence $\sigma_2, \ldots, \sigma_n > 0$ such that, for all $\varepsilon > 0$ and all $2 \leqslant k \leqslant n$,*

$$\kappa_k \leqslant \frac{A}{\sigma_k} + 0.4\varepsilon + \frac{1}{\varepsilon\sigma_k} \sum_{l=1}^{k-1} A_{k,l}\kappa_l. \tag{2.8} \quad \{22\}$$

Then,

$$\kappa_n \leqslant \frac{1}{\sigma_n} \frac{\big(5(A \vee 1) + 2\alpha_n + \alpha'_n\big)\big(2\alpha_n + \alpha'_n\big)}{5\alpha'_n},$$

where

$$\alpha_n = \sup_{2 \leqslant k \leqslant n} \sum_{l=1}^{k-1} \frac{\sigma_k}{\sigma_l} A_{k,l}, \qquad \alpha'_n = \sqrt{2\alpha_n(2\alpha_n + 5(A \vee 1))}.$$

Example 2.1. Let $W_n = n^{-1/2} \sum_{i=1}^n X_i$ where X_i are i.i.d. with $\mathbb{E}X_i = 0$ and $\mathrm{Var}\, X_i = 1$ and $\mathbb{E}|X_i|^3 = \gamma \geqslant 1$. Let $\kappa_n = d_{\mathrm{K}}\big(\mathscr{L}(W_n), \mathrm{N}(0,1)\big)$. Set $G = -n^{1/2}X_I$ and $W' = W - n^{-1/2}X_I$. Set also $W'' = W'$, $\tilde{D} = D$ and $S = 1$. Hence $D = D' = -n^{-1/2}X_I$. We have

$$r_0 = r_1 = r_2 = r_3 = 0.$$

Furthermore,

$$r_9 \leqslant 6\gamma/\sqrt{n}, \qquad r_{10} \leqslant 3\gamma/\sqrt{n}$$

Note now (c.f. Lemma 6.2)

$$\vartheta_\varepsilon(X_n) = \sup_a \mathbb{P}[a \leqslant W_n \leqslant a + \varepsilon | X_n]$$

$$= \sup_a \mathbb{P}\left[\sqrt{\tfrac{n}{n-1}}a \leqslant W_{n-1} + \sqrt{\tfrac{1}{n-1}}X_n \leqslant \sqrt{\tfrac{n}{n-1}}a + \sqrt{\tfrac{n}{n-1}}\varepsilon + \Big| X_n\right]$$

$$\leqslant \sqrt{\tfrac{n}{2\pi(n-1)}}\varepsilon + 2\kappa_{n-1} \leqslant \varepsilon + 2\kappa_{n-1},$$

hence

$$r_{11}(\varepsilon) \leqslant 2\gamma(\varepsilon + 2\kappa_{n-1})/\sqrt{n}, \qquad r_{12}(\varepsilon) \leqslant \gamma(\varepsilon + 2\kappa_{n-1})/\sqrt{n}.$$

Putting these estimates into Theorem 2.9 we obtain

$$\kappa_k \leqslant \frac{8\gamma}{\sqrt{k}} + 0.4\varepsilon + \frac{5\gamma}{\varepsilon\sqrt{k}}\kappa_{k-1}.$$

We can apply Lemma 2.10 with $A = 8\gamma$, $A_{k,k-1} = 5\gamma$ and $A_{k,l} = 0$ for $l < k - 1$, and $\sigma_k = k^{-1/2}$. We have $\alpha_n = 5\gamma\sqrt{2}$, thus, plugging this into (2.8), $\kappa_n \leqslant 25\gamma/\sqrt{n}$. As this example illustrates, the constants obtained this way are typically not optimal, but nevertheless explicit.

3. COUPLINGS

In this section we present some well-known and some new couplings and show how they can be represented in our general framework. The basis is always the coupling (W, W', G) and throughout this section (with the exception of classic exchangeable pairs) we will only look at cases of actual Stein couplings, that is, where $r_0 = 0$. This implies in particular that $\mathbb{E}W = 0$ (which, nevertheless, has to be assumed explicitly in some cases to make the construction work in the first place). Unless otherwise stated, the variance σ^2 of W is arbitrary, but finite and non-zero. Note that, if (W, W', G) is a Stein coupling, so is $(W/\sigma, W'/\sigma, G/\sigma)$, and hence we will usually omit the standardising constant σ^{-1} for ease of notation. To simplify or optimize the bounds, we sometimes will extend the coupling by different choices of $\tilde{D}$, W'' and S. But, if not otherwise stated, we will make the basic assumption throughout this section that $\tilde{D} = D$, $W'' = W$ and $S = 1$.

We mostly present the construction of the couplings only and not the particular form of the final bounds for the normal approximation. The reason for this is that, once the coupling is constructed, one can directly apply our theorems or corollaries of the main section to obtain the corresponding bounds. Hence, stating them explicitly would be either just repeating known results from the literature or rephrasing the results from the main section.

We need to clarify again that a Stein coupling by itself does by no means imply closeness to normality or imply any convergence. As can be seen from the case of quadratic forms (Section 3.2.3), Stein couplings as defined by (1.8) can also be used for χ^2 approximation.

Let throughout this part $[n] := \{1, 2, \ldots, n\}$ and $[0] := \emptyset$. Let also in general I and J be independent random variables, uniformly distributed on $[n]$ and independent of all else, but we will usually mention this—and deviations from it—explicitly whenever we make use of these random variables.

3.1. **Exchangeable pairs and extensions.** This approach was introduced by Stein in a paper by Diaconis (1977). A systematic exposition was given by Stein (1986).

Construction 1A. *Assume that (W, W') is an exchangeable pair. If, for some constant $\lambda > 0$, we have*

$$\mathbb{E}^W(W' - W) = -\lambda W, \qquad (3.1) \quad \{23\}$$

then $(W, W', \frac{1}{2\lambda}(W' - W))$ is a Stein coupling.

Rinott and Rotar′ (1997) generalised (3.1) to allow for some non-linearity in (3.1); however, the resulting coupling will only be an approximate Stein coupling.

Construction 1B. *Assume that (W, W') is an exchangeable pair where $\mathbb{E}W = 0$ and $\operatorname{Var} W = 1$. Assume that, for some constant $\lambda > 0$, we have*

$$\mathbb{E}^W(W' - W) = -\lambda W + R. \qquad (3.2) \quad \{24\}$$

then, with $G = \frac{1}{2\lambda}(W' - W))$,

$$r_0 \leqslant \lambda^{-1}\mathbb{E}|R|, \qquad |\mathbb{E}(GD) - 1| \leqslant \lambda^{-1}|\mathbb{E}(WR)| \leqslant \lambda^{-1}\sqrt{\operatorname{Var} R}.$$

(note that we use $\operatorname{Var} W = 1$ only to obtain the last two inequalities).

The conditional expectation can of course always be written in the form of (3.2) for any λ. However, we will need $\lambda^{-1}\sqrt{\operatorname{Var} R} \to 0$ to obtain convergent bounds, and in this sense the choice of λ is, at least asymptotically, unique; see the discussion in the introduction of Reinert and Röllin (2009a).

Note that we call this approach 'classic' for this specific choice of G. There are many other ways to construct Stein couplings where (W, W') is exchangeable but G is not a multiple of $W' - W$; we will give such examples later on.

The classic exchangeable pairs approach is frequently used in the literature; see for example Rinott and Rotar′ (1997), Fulman (2004a), Fulman (2004b), Röllin (2007b), Meckes (2008) and others. Generally, one constructs a "natural" exchangeable pair (W', W) and then hopes that (3.2) holds with $R = 0$ or R small enough to yield convergence. However, more often than not, this will not succeed, even for simple examples as the 2-runs examples below illustrates. Based on work by Reinert and Röllin (2009a), we will present in Section 3.1.1 Stein couplings making use of a multivariate extensions of (3.1) which will lead to appropriate modifications of G such that (1.8) holds. In Sections 3.2.4 and 3.4 we will also present two very general couplings that are based on exchangeable pairs, but where G is chosen rather differently.

A few more detailed remarks about this approach are appropriate here. For this specific choice of $G = (W' - W)/2\lambda$, Röllin (2008a) proves that exchangeability is actually not necessary to prove a result such as Theorem 2.5, as long as we have equal marginals $\mathscr{L}(W') = \mathscr{L}(W)$. Röllin (2008a) uses a different way of deducing a Stein identity of the form (1.11). With $F(w) = \int_0^w f(x)dx$ one obtains from Taylor's expansion that

$$F(W') - F(W) = Df(W) + D\int_0^D (1 - s/D)f'(W + s)ds,$$

so that, again with $G = D/2\lambda$, and assuming $\mathscr{L}(W') = \mathscr{L}(W)$,

$$\mathbb{E}\{Gf(W)\} = \mathbb{E}\left\{G \int_0^D (1 - s/D)f'(W + D)ds\right\}, \qquad (3.3) \quad \{25\}$$

which serves as a replacement for (1.11). In contrast, Stein (1986) uses the antisymmetric function approach. If (W, W') is exchangeable then $\mathbb{E}\{(W' - W)(f(W') + f(W))\} = 0$, and it is not difficult to show that

$$\mathbb{E}\{Gf(W)\} = \frac{1}{2}\mathbb{E}\left\{G \int_0^D f'(W + D)ds\right\}. \qquad (3.4) \quad \{26\}$$

Note that this is almost (3.3) except that the factor $(1 - s/D)$ is replaced by $1/2$. Note again that (3.4) is only true under exchangeability whereas (3.3) holds for equal marginals. Surprisingly, better constants can be obtained if (3.3) is used instead of (3.4), although exchangeability is a stronger assumption. Incidentally, in Section 4.4, a coupling is used which is not an exchangeable pair but has equal marginals, however, in the context of a different construction than Construction 1A.

3.1.1. *Multivariate exchangeable pairs.* In Reinert and Röllin (2009a), the classic exchangeable pairs approach was generalised to d-dimensional vectors $W = (W_1, \ldots, W_d)$ and $W' = (W'_1, \ldots, W'_d)$ which satisfy

$$\mathbb{E}^W(W' - W) = -\Lambda W \qquad (3.5) \quad \{27b\}$$

for some invertible $(d \times d)$-matrix Λ. They are able to obtain multivariate normal approximation results in cases where the exchangeable pair of univariate random variables (W_1, W'_1) does not satisfy (3.1), but, using auxiliary random variates, an embedding of that pair into a higher dimensional space satisfies (3.5). However, the transition to higher dimensions comes at the cost of having to impose stronger conditions on the set of test functions. Hence, besides the multivariate approximation, it is therefore still of interest to examine W_1 directly. It turns out that, once the higher dimensional embedding satisfying (3.5) is found, it is easy to construct a Stein coupling from that.

Construction 1C. *Let (W, W') be an exchangeable pair of d-dimensional random vectors satisfying (3.6) for some invertible Λ. Let e_i be the i-th unit vector. Then*

$$\left(W_i, W'_i, \tfrac{1}{2}e_i^t\Lambda^{-1}(W' - W)\right)$$

is a Stein coupling.

Indeed,

$$-\mathbb{E}\{Gf(W_i)\} = -\tfrac{1}{2}\mathbb{E}\{e_i^t\Lambda^{-1}\mathbb{E}^W(W' - W)f(W_i)\}$$
$$= \tfrac{1}{2}\mathbb{E}\{e_i^t\Lambda^{-1}\Lambda W f(W_i)\} = \tfrac{1}{2}\mathbb{E}\{W_i f(W_i)\}.$$

and, using exchangeability, the corresponding result for $\mathbb{E}\{Gf(W'_i)\}$ can be obtained in the same way. Hence, every multidimensional exchangeable pair (W, W') satisfying (3.5) gives rise to a univariate Stein coupling for each individual coordinate.

Let us consider the case of 2-runs on a circle. To this end, let $\xi_1, \ldots, \xi_n$ be a sequence of independent $\mathrm{Be}(p)$ distributed random variables. Let $V =$

$\sum_{i=1}^{n}(\xi_i\xi_{i+1}-p^2)$ be the centered number of 2-runs, where we put $\xi_{n+1}=\xi_1$ (hence 'circle'). Consider now the following coupling. With $\xi'_1,\dots,\xi'_n$ being independent copies of $\xi_1,\dots,\xi_n$, let $V'=V-\xi_{I-1}\xi_I-\xi_I\xi_{I+1}+\xi_{I-1}\xi'_I+\xi'_I\xi_{I+1}$, where I is uniformly distributed on $[n]$ and independent of all else. It is easy to see that (V,V') is an exchangeable pair and that

$$\mathbb{E}^V(V'-V)=-\frac{2}{n}V+\frac{2p}{n}\sum_{i=1}^{n}(\xi_i-p). \qquad (3.6) \quad \{27\}$$

Even in this very simple example, the linearity condition (3.2) cannot be obtained with the above natural coupling. Based on the same exchangeable pair, Reinert and Röllin (2009a) use the embedding method to circumvent this problem. To this end we introduce the auxiliary statistic $U=\sum_{i=1}^{n}(\xi_i-p)$ and define $U'=U-\xi_I+\xi'_I$. Condition (3.5) is now satisfied for $W=(U,V)$, $W'=(U',V')$ and

$$\Lambda=\frac{1}{n}\begin{bmatrix} 1 & 0 \\ -2p & 2 \end{bmatrix}.$$

The inverse of Λ is

$$\Lambda^{-1}=n\begin{bmatrix} 1 & 0 \\ p & 1/2 \end{bmatrix},$$

hence Construction 1C yields

$$\begin{aligned} G &= \frac{n}{2}\big(p(U'-U)+\tfrac{1}{2}(V'-V)\big) \\ &= \frac{n}{2}\big(p\xi'_I-p\xi_I+\tfrac{1}{2}\xi_{I-1}\xi'_I+\tfrac{1}{2}\xi'_I\xi_{I+1}-\tfrac{1}{2}\xi_{I-1}\xi_I-\tfrac{1}{2}\xi_I\xi_{I+1}\big) \end{aligned} \qquad (3.7) \quad \{27\text{d}\}$$

so that (V,V',G) is a Stein coupling. Exploiting some specific properties in this example, we may also choose

$$G=\frac{n}{2}\big(p\xi'_I-p\xi_I+\xi'_I\xi_{I+1}-\xi_I\xi_{I+1}\big) \qquad (3.8) \quad \{27\text{c}\}$$

to obtain a somewhat simpler Stein coupling, but the similarity between (3.7) and (3.8) is apparent. See Reinert and Röllin (2009a), Reinert and Röllin (2009b) and Ghosh (2010) for further examples of multivariate exchangeable pair couplings.

3.1.2. *Finding W for a given coupling and a given G.* In some cases it may not be clear from the beginning how to choose the main random variable of interest W. Consider the Curie-Weiss model of ferromagnetic interaction. With $\beta\geqslant 0$ being the *inverse temperature* and $h\in\mathbb{R}$ the *external field* on the state space $\{-1,1\}^n$, we define the probabilities for each $\sigma=(\sigma_1,\dots,\sigma_n)\in\{-1,1\}^n$ by the Gibbs measure

$$\mathbb{P}[\{\sigma\}]=Z^{-1}\exp\bigg\{\frac{\beta}{n}\sum_{i<j}\sigma_i\sigma_j+h\sum_i\sigma_i\bigg\} \qquad (3.9) \quad \{28\}$$

where $Z=Z(\beta,h,n)$ is the partition function to make the probabilities sum up to 1. A quantity of interest is the magnetization $m(\sigma)=n^{-1}\sum_i\sigma_i\in[-1,1]$ of the system. However, in the low-temperature regime the system will exhibit *spontaneous magnetization* so that we may not be interested in $m(\sigma)$ itself if σ is drawn at random according to (3.9) but in $m(\sigma)$ *relative to*

its corresponding magnetization. To find a suitable correction term (which shall serve here as an illustrative example only), Chatterjee (2007) proposes the following construction.

Construction 1D. *Let (σ, σ') be an exchangeable pair on some measure space and let $\varphi(\sigma, \sigma')$ be an anti-symmetric function. Let $G = -\varphi(\sigma, \sigma')/2$, $W = W(\sigma) = \mathbb{E}^{\sigma}\varphi(\sigma, \sigma')$ and $W' = W(\sigma') = \mathbb{E}^{\sigma'}\varphi(\sigma', \sigma)$. Then this defines a Stein coupling.*

Indeed,

$$-\mathbb{E}\big\{Gf(W)\big\} = \tfrac{1}{2}\mathbb{E}\big\{\varphi(\sigma, \sigma')f(W)\big\} = \tfrac{1}{2}\mathbb{E}\big\{Wf(W)\big\}$$

and

$$\begin{aligned}
\mathbb{E}\big\{Gf(W')\big\} &= -\tfrac{1}{2}\mathbb{E}\big\{\varphi(\sigma, \sigma')f(W(\sigma'))\big\} \\
&= -\tfrac{1}{2}\mathbb{E}\big\{\varphi(\sigma', \sigma)f(W(\sigma))\big\} \\
&= \tfrac{1}{2}\mathbb{E}\big\{\varphi(\sigma, \sigma')f(W)\big\} = \tfrac{1}{2}\mathbb{E}\big\{Wf(W)\big\},
\end{aligned}$$

which proves (1.8).

Let us apply this to the Curie-Weiss model. First, given σ is drawn from (3.9), we define σ' by choosing a site I uniformly at random and then we resample this site according to the conditional distribution $\mathscr{L}(\sigma_I \,|\, \sigma_j, j \neq I)$, giving a new σ'_I, but leaving all the other sites untouched. Now we set $\varphi(\sigma, \sigma') = n(m(\sigma) - m(\sigma')) = \sigma_I - \sigma'_I$. It is not difficult to show that

$$\mathbb{E}^{\sigma}\varphi(\sigma, \sigma') = m(\sigma) - \frac{1}{n}\sum_i \tanh(\beta m_i(\sigma) + \beta h),$$

where $m_i(\sigma) = \frac{1}{n}\sum_{j \neq i}\sigma_j$. Hence, we let $G = -(\sigma_I - \sigma'_I)/2$,

$$W = W(\sigma) = m(\sigma) - \frac{1}{n}\sum_i \tanh(\beta m_i(\sigma) + \beta h),$$

and $W' := W(\sigma')$. As

$$\left|\tanh(\beta m(\sigma) + \beta h) - \frac{1}{n}\sum_i \tanh(\beta m_i(\sigma) + \beta h)\right| \leqslant \frac{\beta}{n},$$

we may alternatively choose $W = m(\sigma) - \tanh(\beta m(\sigma) + \beta h)$, in which case (1.8) is not satisfied anymore, but we still have $r_0 \leqslant \beta/n$. The key here is to find $\varphi(\sigma, \sigma')$ such that $\mathbb{E}^{\sigma}\varphi(\sigma, \sigma')$ yields 'something interesting'. In Chatterjee (2007) this construction is used to prove concentration of measure results for such W.

3.1.3. *Finding an antisymmetric G through a Poisson equation.* Assume now that (X, X') is an exchangeable pair on some space $\mathcal{X}$ and $W = \varphi(X)$ and $W' = \varphi(X')$ for some functional $\varphi : \mathcal{X} \to \mathbb{R}$ with $\mathbb{E}\varphi(X) = 0$. Chatterjee (2005, Section 4.1) proposes a general approach to find G of a special form. Let $G(X', X) = \tfrac{1}{2}(\psi(X') - \psi(X))$ for some unknown functional $\psi : \mathcal{X} \to \mathbb{R}$ (in fact, *any* anti-symmetric function can be written in this form; see Stein (1986)). Using exchangeability, it is not difficult to see that (1.8) is satisfied if

$$\psi(x) - P\psi(x) = \varphi(x)$$

for every $x \in \mathcal{X}$, which we recognize as a Poisson equation with kernel $P\psi(x) := \mathbb{E}^{X=x}\psi(X')$ for given φ and unknown ψ. A general (formal) solution is $\psi(x) = \sum_{k=0}^{\infty} P^k\varphi(x)$. We have the following.

Construction 1E. *Let (X, X') be an exchangeable pair on a measure space $\mathcal{X}$ and let $\varphi : \mathcal{X} \to \mathbb{R}$ be a measurable function such that $\mathbb{E}\varphi(X) = 0$; define P as above. If there is a constant $C > 0$ such that*

$$\sum_{k=0}^{\infty} |P^k\varphi(x) - P^k\varphi(y)| \leqslant C \qquad (3.10) \quad \{29\}$$

for every $x, y \in \mathcal{X}$, then

$$(W, W', G) = \Big(\varphi(X), \varphi(X'), \frac{1}{2}\sum_{k=0}^{\infty}\big(P^k\varphi(X) - P^k\varphi(X')\big)\Big)$$

is a Stein coupling.

Note that boundedness $|G| \leqslant C/2$ is built in through (3.10) so that this construction is a natural candidate for Theorem 2.5. We can give a more constructive version of this coupling.

Construction 1F. *Assume that (X, X'), φ and P are as in Construction 1E. Assume that we have two Markov chains $(X_n)_{n \geqslant 0}$ and $(X'_n)_{n \geqslant 0}$ with the transition dynamics given by P, and also $X_0 = X$ and $X'_0 = X'$. Assume further that, for all n,*

$$\mathscr{L}\big(X_n \big| X, X'\big) = \mathscr{L}\big(X_n \big| X\big), \quad \mathscr{L}\big(X'_n \big| X, X'\big) = \mathscr{L}\big(X'_n \big| X'\big). \qquad (3.11)$$

Let now $T = \inf\{n > 0 \mid X_n = X'_n\}$ be the coupling time of the two chains and assume that $T < \infty$ almost surely. If, given T, I is uniformly distributed on $\{0, 1, 2, \ldots, T-1\}$, then

$$(W, W', G) = \big(\varphi(X), \varphi(X'), \tfrac{1}{2}T\big(\varphi(X_I) - \varphi(X'_I)\big)\big) \qquad (3.12) \quad \{30\}$$

is a Stein coupling.

Indeed, from

$$\mathbb{E}\big\{\varphi(X_k)f(W)\big\} = \mathbb{E}\big\{P^k(X)f(W)\big\}$$

and

$$\mathbb{E}\big\{\varphi(X'_k)f(W)\big\} = \mathbb{E}\big\{f(W)\mathbb{E}^{X,X'}\varphi(X'_k)\big\} = \mathbb{E}\big\{f(W)P^k\varphi(X')\big\},$$

we easily obtain

$$\mathbb{E}\big\{T\big(\varphi(X_I) - \varphi(X'_I)\big)f(W)\big\} = \mathbb{E}\sum_{k=0}^{T-1}\big(\varphi(X_k) - \varphi(X'_k)\big)f(W)$$

$$= \sum_{k=0}^{\infty}\mathbb{E}\big\{\big(\varphi(X_k) - \varphi(X'_k)\big)f(W)\big\}$$

$$= \sum_{k=0}^{\infty}\mathbb{E}\big\{\big(P^k\varphi(X) - P^k\varphi(X')\big)f(W)\big\},$$

and, similarly,

$$\mathbb{E}\{T(\varphi(X_I) - \varphi(X'_I))f(W')\} = \sum_{k=0}^{\infty} \mathbb{E}\{(P^k\varphi(X) - P^k\varphi(X'))f(W')\}$$
$$= -\sum_{k=0}^{\infty} \mathbb{E}\{(P^k\varphi(X) - P^k\varphi(X'))f(W)\},$$

where the second step uses exchangeability. Hence, it follows from Construction 1E that (3.12) is a Stein coupling; see Chatterjee (2005, Section 4.1) for more details, and see also Makowski and Shwartz (1994) on general theory about Poisson equations.

3.2. **Local dependence and related couplings.** This is one of the earliest versions of Stein's method. Let in what follows I be uniformly distributed on $[n]$, independent of all else.

Construction 2A. *Let $W = \sum_{i=1}^n X_i$ with $\mathbb{E}X_i = 0$. For each i, let W'_i be such that*

$$\mathbb{E}(X_i \mid W'_i) = 0. \qquad (3.13) \quad \{31\}$$

Then, $(W, W', G) = (W, W'_I, -nX_I)$ is a Stein coupling.

To see this we have on one hand

$$-\mathbb{E}\{Gf(W)\} = \sum_{i=1}^n \mathbb{E}\{X_i f(W)\} = \mathbb{E}\{Wf(W)\}, \qquad (3.14) \quad \{32\}$$

and on the other hand

$$\mathbb{E}\{Gf(W')\} = -\sum_{i=1}^n \mathbb{E}\{X_i f(W'_i)\} = 0,$$

due to (3.13); hence (1.8) is satisfied.

The choice $G = -nX_I$ was first considered by Stein (1972) for m-dependent sequences, however this G has broader applications. We now discuss some more detailed constructions of W'_i below.

3.2.1. *Local dependence.* Local dependence was extensively studied in Chen and Shao (2004) under various dependence settings, but of course this approach goes back to Stein (1972) and Chen (1975); a version for discrete random variables is given by Röllin (2008b). We can use the simplest form as a starting point

Construction 2B. *Assume that W and G are as in Construction 2A. Assume in addition that, for each $i \in [n]$, there is $A_i \subset [n]$ such that X_i and $(X_j)_{j \in A_i^c}$ are independent. Then, with $W'_i = W - \sum_{j \in A_i} X_j$, (3.13) is satisfied.*

This first-order dependence is usually referred to as (LD1) and is enough to obtain a Stein coupling. However, it is possible to extend this coupling.

Construction 2C. *Assume that W and G and W_i are as in Constructions 2A and 2B and that $\mathrm{Var}\,W = 1$. Assume in addition that there is $B_i \subset [n]$ such that $A_i \subset B_i$ and $(X_j)_{j \in A_i}$ and $(X_j)_{j \in B_i^c}$ are independent. Define $W''_i = W - \sum_{j \in B_i} X_j$; then $W'' := W''_I$ is independent of GD and hence $r_3 = 0$.*

The conditions of Constructions 2B and 2C together are referred to as (LD2).

3.2.2. *Decomposable random variables.* This version of local dependence was popularized by Barbour et al. (1989) for smooth test functions, by Raič (2004) for Kolmogorov distance and by Röllin (2008b) for total variation approximation of discrete random variables. It uses a refined version of the concept of second-order neighborhood and makes use of non-trivial $\tilde{D}$ and S.

Construction 2D. *Assume that W and G and W_i are as in Constructions 2A and 2B. Assume in addition that, for each i and for each $j \in A_i$, there is $B_{i,j} \subset [n]$ such that $A_i \subset B_{i,j}$ and such that (X_i, X_j) is independent of $(X_j)_{j \in B_{i,j}^c}$. Let $K_i = |A_i|$ and define $\tilde{D} = K_I X_J$ where, given I, J is uniformly distributed on $[K_I]$, but independent of all else. Then, $\mathbb{E}^X(G\tilde{D}) = \mathbb{E}^X(GD)$, hence $r_1 = 0$. Let $S = nK_I \sigma_{I,J}$ where $\sigma_{i,j} = \mathbb{E}(X_i X_j)$, then $r_2 = 0$. Define $W''_{i,j} = W - \sum_{k \in B_{i,j}} X_k$; then $W'' := W''_{I,J}$ is independent of $G\tilde{D}$ and S and hence $r_3 = 0$.*

Hence, if we can choose $B_{i,j}$ such that $B_{i,j} \subsetneq B_i$, where B_i is as for the standard (LD2) local dependence setting from Construction 2C, we should be able to improve our bounds, as $W'' - W = \sum_{k \in B_{I,J}} X_k$ contains fewer summands as compared to $\sum_{k \in B_I} X_k$ from Construction 2C.

Note that, under third moment conditions, Barbour et al. (1989) obtain a Wasserstein bound of order

$$\sum_{i=1}^n \mathbb{E}|X_i Z_i^2| + \sum_{i=1}^n \sum_{j \in A_i} \big(\mathbb{E}|X_i X_j V_{i,j}| + |\mathbb{E}(X_i X_j)|\mathbb{E}|Z_i + V_{i,j}|\big) \qquad (3.15) \quad \{33\}$$

with

$$Z_i := \sum_{j \in A_i} X_k, \qquad V_{i,j} := \sum_{k \in B_{i,j} \setminus A_i} X_k,$$

(note that in Barbour et al. (1989) the coarser expression $\mathbb{E}|X_i X_j|$ is used, but it is easy to see that this can be sharpened to $|\mathbb{E}(X_i X_j)|$. Using Corollary 2.3, we obtain an order of

$$\sum_{i=1}^n \mathbb{E}|X_i Z_i^2| + \sum_{i=1}^n \sum_{j \in A_i} \big(\mathbb{E}|X_i X_j (Z_i + V_{i,j})| + |\mathbb{E}(X_i X_j)|\mathbb{E}|Z_i + V_{i,j}|\big). \quad (3.16) \quad \{34\}$$

In most cases we can expect that these two bounds will yield similar results: Indeed, a useful upper bound on both estimates is

$$\sum_{i=1}^n \sum_{j \in A_i} \sum_{k \in B_{i,j}} \big(\mathbb{E}|X_i X_j X_k| + |\mathbb{E}(X_i X_j)|\mathbb{E}|X_k|\big)$$

up to constants.

Consider the von Mises statistics as an example, where, for independent random variables $X_1, \ldots, X_n$, we have $W = \sum_{p,q} \varphi_{p,q}(X_p, X_q)$ for some functionals $\varphi_{p,q}$. Clearly, we can choose $A_{(p,q)} = \{(k,l) : k = q \text{ or } l = q\}$ as first-order neighborhood. However, in the standard local approach framework, we would need to let $B_{(p,q)} = [n] \times [n]$ for (LD2) so that $W'' = 0$ and hence $|D'| = |W|$ which would not yield useful bounds. In the refined setting we can choose, for every $(p', q') \in A_{(p,q)}$, the set $B_{(p,q),(p',q')} := \{(k,l) :$

$k \in \{p, p'\}$ or $l \in \{q, q'\}\}$, so that now $|B_{(p,q),(p',q')}|$ is only of order n. Of course, it will depend on the concrete choice of functionals $\varphi_{p,q}$ whether normal approximation is appropriate at all; see Barbour et al. (1989) for applications to random graph related statistics.

3.2.3. *Special case: quadratic forms.* Let $\xi_1, \ldots, \xi_n$ be independent, centered random variables with unit variance. Let $A = (a_{ij})$ be a real symmetric $(n \times n)$-matrix. Let $W = \left(\sum_{i,j} a_{ij} \xi_i \xi_j - \sum_i a_{ii} \right)$. It would be straightforward to use the above method of decomposable random variables in this situation. However, due to the multiplicative structure (or, in U-statistics language, because the kernel $\varphi_{i,j}(x, y) = a_{ij} xy$ is degenerate for centered random variables) there is an interesting alternative.

Construction 2E. *Let W be as above. Let also $Y_i := \sum_j a_{ij} \xi_j$, $G = -n(\xi_I Y_I - a_{II})$ and $W' = W - (2\xi_I Y_I - a_{II} \xi_I^2)$. Then this defines a Stein coupling.*

It is not difficult to see that (1.8) holds. Again, it depends on the matrix A whether we can expect normal like behaviour of W or not: essentially this coupling was used by Luk (1994) in the context of χ^2-approximation for the case where all the entries of A are 1, corresponding to the square of a sum of random variables.

3.2.4. *Local exchangeable randomization.* This coupling was proposed in Reinert (1998). Its use was limited by the fact that, if the classic exchangeable pairs approach (as discussed in Subsection 3.1) is used along with this coupling, the linearity condition (3.2) will in general not be satisfied with R small enough, but with the choice $G = -nX_I$, we can now handle this coupling. However, some care is needed.

Construction 2F. *Let W and G be as in Construction 2A. Let $(X'_{i,j})_{i,j \in [n]}$ be a collection of random variables, such that, with $W'_i = \sum_{j=1}^{n} X'_{i,j}$, we have that*

$$(i) \quad \text{for each } i, X'_{i,i} \text{ is independent of } W, \qquad (3.17) \quad \{37\}$$

$$(ii) \quad \text{for each } i, \big((X_k)_k, (X'_{i,k})_k\big) \text{ is exchangeable.} \qquad (3.18) \quad \{38\}$$

Then (3.13) is satisfied.

It is often not too difficult to construct $(X'_{i,j})_j$ for a given i such that $\mathscr{L}(W'_i) = \mathscr{L}(W)$ and such that $X'_{i,i}$ is independent of W and hence $\mathbb{E}^W X'_{i,i} = 0$. However, it is important to note that this does not suffice as we ultimately need $\mathbb{E}^{W'_i} X_i = 0$, which is, however, guaranteed under the additional Condition (3.18).

In Reinert (1998), it was incorrectly deduced from (3.17) and the property

$$\mathscr{L}(X'_{i,j}, j \neq i \,|\, X'_{i,i} = x) = \mathscr{L}(X_j, j \neq i \,|\, X_i = x) \qquad (3.19) \quad \{39\}$$

that (W, W'_i) is exchangeable. It is not difficult to find examples for which (W, W'_i) is not exchangeable, but (3.17) and (3.19) are still true; see Remark 4.2.

3.3. **Size-biasing.** This approach was introduced in Baldi, Rinott, and Stein (1989) and further explored in Goldstein and Rinott (1996), Dembo and Rinott (1996), Goldstein and Penrose (to appear) and others.

Construction 3A. *Let V be a non-negative random variable with $\mathbb{E}V = \mu > 0$. Let V^s have the* size-biased *distribution of V, that is, for all bounded f,*

$$\mathbb{E}\{Vf(V)\} = \mu\mathbb{E}f(V^s). \qquad (3.20) \quad \{40\}$$

Then

$$(W, W', G) = (V - \mu, V^s - \mu, \mu)$$

is a Stein coupling.

Using (3.20), we obtain

$$\begin{aligned}
\mathbb{E}\{Gf(W') - Gf(W)\} &= \mathbb{E}\{\mu f(V^s - \mu) - \mu f(V - \mu)\} \\
&= \mathbb{E}\{Vf(V - \mu) - \mu f(V - \mu)\} \\
&= \mathbb{E}\{Wf(W)\},
\end{aligned}$$

so that (1.8) is satisfied, indeed.

One of the advantages of this approach is apparent if bounds for the Kolmogorov metric are to be obtained. In the light of Theorem 2.1, we see that G/σ is already bounded by $\alpha = \mu/\sigma$, so that we only need to concentrate on finding a bounded coupling (W, W'); see Goldstein and Penrose (to appear) for such a coupling in the context of coverage problems.

3.4. **Interpolation to independence.** For this coupling the key idea is to construct a sequence of random variables that 'interpolates' between W and an independent copy of W by means of small perturbations. A special case of this coupling was introduced by Chatterjee (2008). The construction has apparent similarities to Lindeberg's telescoping sum in his prove of the CLT for sums of independent random variables. Let in the following construction I be uniformly distributed on $[n]$ and independent of all else.

Construction 4A. *Assume $\mathbb{E}W = 0$. Assume that for each $i \in [n]$ we have a W_i' which is close to W. Assume that there is a sequence of random variables $V_0, V_1, \ldots, V_n$ such that $\mathbb{E}^W V_0 = W$ and such that V_n is independent of V_0 and assume that, for every $i \in [n]$,*

$$\mathscr{L}\big[(W, V_{i-1}), (W_i', V_i)\big] = \mathscr{L}\big[(W_i', V_i), (W, V_{i-1})\big] \qquad (3.21) \quad \{42\}$$

for every $i \in [n]$. Then

$$(W, W', G) = \big(W, W_I', \tfrac{n}{2}(V_I - V_{I-1})\big) \qquad (3.22) \quad \{43\}$$

is a Stein coupling.

Note that (3.21) implies in particular that (W, W_i') is an exchangeable pair for each i and also that $\mathscr{L}(V_i) = \mathscr{L}(V_0)$ for all i by induction. We have

$$\begin{aligned}
\mathbb{E}\{Gf(W)\} &= \tfrac{1}{2}\mathbb{E}\sum_{i=1}^{n}(V_i - V_{i-1})f(W) \\
&= \tfrac{1}{2}\mathbb{E}\{(V_n - V_0)f(W)\} \\
&= -\tfrac{1}{2}\mathbb{E}\{Wf(W)\},
\end{aligned}$$

due to the independence assumption, and, due to (3.21),

$$\mathbb{E}\{Gf(W')\} = \frac{1}{2} \sum_{i=1}^{n} \mathbb{E}\{(V_i - V_{i-1})f(W_i')\}$$

$$= \frac{1}{2} \sum_{i=1}^{n} \mathbb{E}\{(V_{i-1} - V_i)f(W)\}$$

$$= -\frac{1}{2} \sum_{i=1}^{n} \mathbb{E}\{(V_i - V_{i-1})f(W)\}$$

$$= -\mathbb{E}\{Gf(W)\} = \tfrac{1}{2}\mathbb{E}\{Wf(W)\}.$$

Hence, (3.22) is a Stein coupling, indeed.

3.4.1. *Functionals of independent random variables.* A specific version of this coupling was used by Chatterjee (2008) for functionals of independent random variables. We give a simpler version first and discuss then the (implicitly used) coupling of Chatterjee (2008).

Construction 4B. *Let $X = (X_1, \ldots, X_n)$ be a collection of independent random variables and let $W = F(X)$ be any functional of X such that $\mathbb{E}F(X) = 0$. Let $X' = (X_1', \ldots, X_n')$ be an independent copy of X and define for all subsets $A \subset [n]$ the vectors $X^A = (X_1^A, \ldots, X_n^A)$ by*

$$X_i^A = \begin{cases} X_i' & \text{if } i \in A, \\ X_i & \text{if } i \notin A, \end{cases}$$

that is, X^A is simply X but with all X_i replaced by X_i' for which $i \in A$; define also $W_A' = F(X^A)$. Let $W_i' := W_{\{i\}}'$ for each $i \in [n]$ and $V_i := W_{[i]}'$ for each $i \in [n] \cup \{0\}$. Then the conditions of Construction 4A are satisfied.

Clearly, V_n is independent of V_0 and it is not difficult to see that (3.21) holds. The interpolating sequence is therefore constructed simply by replacing the X_i by X_i' in increasing order. For this coupling to be useful we would typically need that F is not too sensitive to changes in the individual coordinates.

The implicit coupling used by Chatterjee (2008) is different in the sense that, instead of using a fixed order in which the X_i are replaced, a random order is used.

Construction 4C. *Assume that W, F, X and X' are as in 4B. Let Π be a uniformly drawn random permutation of length n, independent of everything else. For any permutation π we denote by $\pi(A)$ simply the image of A with respect to π. Define now $W_i' := W_{\{\Pi(i)\}}'$ and $V_i := W_{\Pi([i])}'$. Then the conditions of Construction 4A are satisfied.*

Exchangeability (3.21) follows from Construction 4B by conditioning on Π. Let us now prove that Construction 4C indeed leads to the representation used by Chatterjee (2008). Clearly, $G = \frac{1}{2n}(W_{\Pi([I])}' - W_{\Pi([I-1])}')$, thus

$$\mathbb{E}^{X,X'}(GD) = \frac{1}{2n!} \sum_{i=1}^{n} \sum_{\pi} (W_{\pi([i])}' - W_{\pi([i-1])}')(W - W_{\{\pi(i)\}}').$$

We re-write the sum over all permutation as a sum over all possible subsets induced by $\pi([i-1])$, that is, all possible subsets $A \subset [n]$ with $|A| = i-1$, and over all possible values of $\pi(i)$ which range over $[n] \setminus A$. Taking into account multiplicities from all possible permutations within the sets $\pi([i-1])$ and $\pi([n] \setminus [i])$ we obtain

$$\mathbb{E}^{X,X'}(GD)$$

$$= \frac{1}{2n!} \sum_{i=1}^{n} \sum_{\substack{A \subset [n], \\ |A|=i-1}} \sum_{j \notin A} |A|!(n-|A|-1)!(W'_{A \cup \{j\}} - W'_A)(W - W'_{\{j\}})$$

$$= \frac{1}{2} \sum_{A \subsetneq [n]} \sum_{j \notin A} \frac{1}{\binom{n}{|A|}(n-|A|)}(W'_{A \cup \{j\}} - W'_A)(W - W'_{\{j\}}),$$

which is exactly the expression used by Chatterjee (2008, Eq. (1)).

3.5. Local symmetry.
An instance of this coupling was used by Chen (1998) for sums of independent random variables.

Construction 5A. *Assume that W, W', G_α and G_β are random variables such that*

$$\mathbb{E}\{G_\alpha f(W')\} = \mathbb{E}\{G_\beta f(W')\}, \qquad (3.23) \quad \{45\}$$

$$\mathbb{E}\{G_\alpha f(W)\} = \mathbb{E}\{W f(W)\}, \qquad (3.24) \quad \{46\}$$

$$\mathbb{E}\{G_\beta f(W)\} = 0, \qquad (3.25) \quad \{47\}$$

for all f for which the expectations exist. Then, $(W, W', G_\beta - G_\alpha)$ is a Stein coupling.

Indeed, using (3.23) for the first equality and then (3.24) and (3.25),

$$\mathbb{E}\{(G_\beta - G_\alpha)(f(W') - f(W))\} = \mathbb{E}\{(G_\alpha - G_\beta)f(W)\} = \mathbb{E}\{W f(W)\}.$$

Note that we refer to Condition (3.23) as *local symmetry* due to the following example.

Let $X = (X_1, \ldots, X_n)$ be a sequence of centered independent random variables and let X' be an independent copy of X. Let $W = \sum_i X_i$ and assume that $\operatorname{Var} W = 1$. Define $G_\alpha = X_I$ and $W' = W + X'_I$ (we 'duplicate' a small part of W). Define also $G_\beta = X'_I$. Then it is not difficult to verify Conditions (3.23) and (3.24). Identity (3.23) is due to the symmetry of G_α and G_β relative to W', which is the crucial aspect of the construction.

3.6. Abstract approaches.
One might wonder if, for a given arbitrary coupling (W, W'), one can always find G to make (W, W', G) a Stein coupling.

Construction 6A. *Let (W, W') be a pair of integrable random variables. Let $\mathcal{F}$ and $\mathcal{F}'$ be two σ-algebras with $\sigma(W) \subset \mathcal{F}$ and $\sigma(W') \subset \mathcal{F}'$. Let V be a random variable such that*

$$\mathbb{E}^W V = W. \qquad (3.26) \quad \{47b\}$$

Define (formally) the random variable

$$G = -V + \mathbb{E}(V|\mathcal{F}') - \mathbb{E}(\mathbb{E}(V|\mathcal{F}')|\mathcal{F}) + \mathbb{E}(\mathbb{E}(\mathbb{E}(V|\mathcal{F}')|\mathcal{F})|\mathcal{F}') - \ldots.$$

	Construction	$W' \overset{\mathscr{D}}{=} W$	(W', W) exchangeable	$W'' \overset{\mathscr{D}}{=} W$
Hoeffding (Var. 1)	1A (p. 17)	yes	yes	–
Hoeffding (Var. 2)	2F (p. 24)	yes	yes	–
Hoeffding (Var. 3)	5A (p. 27)	no	no	yes
Occupancy	2F (p. 24)	yes	yes	yes
Neighbourhood	2A (p. 22)	no	no	no
Random graphs	2A (p. 22)	yes	no	yes

TABLE 1. Overview over the couplings used in the different applications along with some interesting properties of the involved random variables.

If the above sequence converges absolutely almost surely, then (W, W', G) is a Stein coupling.

To see this, first condition G on $\mathcal{F}'$ which yields $\mathbb{E}(G|\mathcal{F}') = 0$. On the other hand, $\mathbb{E}(G|\mathcal{F}) = -\mathbb{E}(V|\mathcal{F})$ hence $\mathbb{E}^W G = -W$ and (1.8) follows. Note that (3.26) is not to be confused with the usual linearity condition (3.1)—we can always take $V = W$ to satisfy (3.26).

Consider the example $W = \sum_{i=1}^{n} X_i$, a sum of independent, centered random variables. With I independent and uniformly distributed on $[n]$, let $W' = W - X_I$. Take $V = W$ and note that

$$\mathbb{E}(W|W') = \mathbb{E}(W' + X_I|W') = W', \qquad \mathbb{E}(W'|W) = (1 - \tfrac{1}{n})W.$$

Hence,

$$-G = W - W' + (1 - \tfrac{1}{n})W - (1 - \tfrac{1}{n})W' + (1 - \tfrac{1}{n})^2 W - (1 - \tfrac{1}{n})^2 W' + \dots$$
$$= X_I + (1 - \tfrac{1}{n})X_I + (1 - \tfrac{1}{n})^2 X_I + \dots = n X_I.$$

Alternatively, chosing $V = n X_I$ yields the same result directly, as $\mathbb{E}^{W'} V = 0$ and hence $G = -V = -n X_I$.

4. APPLICATIONS

In this section we give some applications of the main theorems and corollaries using different couplings from Section 3. Table 1 gives an overview over the different couplings we use in this section along with some important characteristics of the involved random variables.

4.1. **Hoeffding's combinatorial statistic.** Let $a_{i,j}$, $1 \leqslant i, j \leqslant n$, be real numbers such that $\sum_{k=1}^{n} a_{i,k} = \sum_{k=1}^{n} a_{k,j} = 0$ and $\frac{1}{n-1} \sum_{i,j} a_{i,j}^2 = 1$. Let π be a uniformly chosen random permutation of size n and $W = \sum_{i=1}^{n} a_{i,\pi(i)}$. Then it is routine to see that $\mathbb{E}W = 0$ and $\operatorname{Var} W = 1$. Note that, for a Stein coupling (W, W', G), unit variance of W implies $\mathbb{E}(GD) = 1$. Let in what follows I_1 and I_2 be independent and uniformly chosen random numbers from $[n]$.

Variant 1 (c.f. Construction 1A). Define $\pi' = \pi \circ (I_1\, I_2)$ and $W' = \sum_{i=1}^{n} a_{i,\pi'(i)}$. Then (W, W') is a classical exchangeable pair, i.e. (3.1) holds with $\lambda = 2/n$, or, equivalently, $G = \frac{n}{4}(W' - W) = \frac{n}{4}(a_{I_1,\pi(I_2)} + a_{I_2,\pi(I_1)} - a_{I_1,\pi(I_1)} - a_{I_2,\pi(I_2)})$ makes (W, W', G) a Stein coupling.

Variant 2 (c.f. Construction 2F). Define W' as in Variant 1. With $G = -na_{I_1,\pi(I_1)}$, (W, W', G) is also a Stein coupling. This coupling is the (implicit) basis for the construction in Ho and Chen (1978) and Bolthausen (1984).

In both of the previous variants, our W' is defined with respect to a perturbation π' of π. Thus, W' can be seen as an instance of the *replacement perturbation* from the introduction. This comes at the cost of D having four terms. One may wonder whether a *deletion* construction is possible, where W' is defined by just 'removing a random small part' of W (see Section 3.2.1). This is possible, indeed, so that we do not need to go through constructing π'. Despite the fact that the following construction is very simple, it has gone unnoticed in the literature so far.

Variant 3 (c.f. Construction 5A). Define

$$W' = W - \begin{cases} (a_{I_1,\pi(I_1)} + a_{I_2,\pi(I_2)}) & \text{if } I_1 \neq I_2, \\ a_{I_1,\pi(I_1)} & \text{if } I_1 = I_2. \end{cases}$$

Let $G = n(a_{I_1,\pi(I_2)} - a_{I_1,\pi(I_1)})$; then (W, W', G) is a Stein coupling. To see this, note first that $\sigma(W') \subset \mathcal{F} := \sigma\big(I_1, I_2, (\pi(i); i \neq I_1, I_2)\big)$. Now, if $I_1 \neq I_2$, the conditional distributions $\mathscr{L}(\pi(I_1)|\mathcal{F})$ and $\mathscr{L}(\pi(I_2)|\mathcal{F})$ are equal and assign probability $1/2$ to each of the points in the set $\{\pi(I_1), \pi(I_2)\}$ so that $\mathbb{E}\{Gf(W')\} = 0$. The same arguments from Variant 1 lead to $\mathbb{E}\{Gf(W)\} = -\mathbb{E}\{Wf(W)\}$.

To see the connection with Construction 5A, let $G_\alpha = na_{I_1,\pi(I_1)}$ and $G_\beta = na_{I_1,\pi(I_2)}$; then (3.23)–(3.25) are satisfied.

Let us quickly illustrate how to obtain a bound in terms of

$$\|a\| := \sup_{1 \leqslant i,j \leqslant n} |a_{i,j}|.$$

With the Stein coupling from Variant 3 we have

$$|G| \leqslant 2n\|a\| =: \alpha, \qquad |D| \leqslant 2\|a\| =: \beta.$$

We will make use of an auxiliary variable W'', which can be constructed so that it is independent of $(I_1, I_2, \pi(I_1), \pi(I_2))$ and such that

$$|D'| \leqslant 8\|a\| =: \beta'. \tag{4.1} \quad \{49\}$$

Hence, applying Theorem 2.5 with the above random variables and constants and in addition $\tilde{D} = D$ and $\tilde{\beta} = \beta$ we easily obtain the following result.

Theorem 4.1. *With W and $\|a\|$ as above,*

$$d_{\mathrm{K}}\big(\mathscr{L}(W), \mathrm{N}(0,1)\big) \leqslant 448n\|a\|^3 + 96\|a\|.$$

Proof. We only need the existence of W'' as claimed above; we use the construction of Bolthausen (1984). It turns out that it is more convenient to construct W from W''. Let τ be a uniformly chosen random permutation of

$[n]$ and let $W'' = \sum_i a_{i,\tau(i)}$. Let (I_1, I_2, J_1, J_2) be random variables independent of τ such that (I_1, I_2, J_1) is uniform on $[n]^3$, such that J_2 is uniform on $[n] \setminus \{J_1\}$ if $I_1 \neq I_2$ and such that $J_1 = J_2$ if $I_1 = I_2$. One can now construct a permutation π which again has uniform distribution, and such that

$$\pi(I_1) = J_1, \qquad \pi(I_2) = J_2,$$

and such that τ and π differ in at most four positions. Now $\mathbb{E}^\tau(G_3(W_3' - W)) = \mathbb{E}(G_3(W_3' - W))$ (and hence $r_3 = 0$) follows from the independence assumption between τ and (I_1, I_2, J_1, J_2). Note that the calculations from Variant 3 above still hold, as, given π, (I_1, I_2) is uniformly distributed on $[n]^2$, hence independent of π as required. $\qquad\square$

Remark 4.1. Note that Goldstein (2005), using zero-biasing, obtains

$$d_{\mathrm{K}}\big(\mathscr{L}(W), \mathrm{N}(0,1)\big) \leqslant 1016\|a\| + 768\|a\|^2.$$

4.2. **Functionals in the classic occupancy scheme.** Let m balls be distributed independently of each other into n boxes such that the probability of landing in box i is p_i, where $\sum_{i=1}^n p_i = 1$. The literature on this topic is rich; see for example Johnson and Kotz (1977), Kolchin, Sevast'yanov, and Chistyakov (1978) or Barbour, Holst, and Janson (1992), but also more recent results such as Hwang and Janson (2008) and Barbour (2009) on local limits theorems for infinite number of urns. If ξ_i denotes the number of balls in urn i after distributing the balls, some interesting statistics can be written in the form

$$U = \sum_{i=1}^n h(\xi_i)$$

for functions $h : \mathbb{Z}_+ \to \mathbb{R}$. Examples are

$\quad h(x) = \mathrm{I}[x = k] \qquad\qquad$ "# urns with exactly k balls",

$\quad h(x) = \mathrm{I}[x > m_0] \qquad\qquad$ "# urns exceeding a limit m_0",

$\quad h(x) = \mathrm{I}[x > m_0](x - m_0) \quad$ "# excess balls when urn limit is m_0";

see for example Boutsikas and Koutras (2002).

Let us consider here the more general case

$$U = \sum_{i=1}^n h_i(\xi_i) \qquad\qquad (4.2) \quad \{52\}$$

for functions $h_i : \mathbb{Z}_+ \to \mathbb{R}$, $i = 1, \ldots, n$. Due to the subsequent centering, we may assume without loss of generality that $h_i(0) = 0$ for all i.

Theorem 4.2. *Assume the situation as described above. Let $W = (U - \mu)/\sigma$, where μ and σ^2 are the mean and the variance of U. Define the quantities*

$$\|h\| = \sup_{1 \leqslant i \leqslant n} \|h_i\|, \quad \|\Delta h\| = \sup_{1 \leqslant i \leqslant n} \sup_{j \in \mathbb{Z}_+} \|h_i(j+1) - h_i(j)\|, \quad \bar{p} = \sup_{1 \leqslant i \leqslant n} p_i,$$

and assume that $\|h\| \geqslant 1$ and $\|\Delta h\| \geqslant 1$. Then, if

$$1 + 10m\bar{p} \leqslant 4\ln(n\|h\|) \leqslant \frac{1}{(2\bar{p})^{1/2}}, \qquad\qquad (4.3) \quad \{53\}$$

we have

$$d_{\mathrm{K}}\big(\mathscr{L}(W), \mathrm{N}(0,1)\big)$$
$$\leqslant \frac{409600 n\|\Delta h\|^3 \ln(n\|h\|)^6(1+\sigma^2/n)}{\sigma^3} + \frac{3888 \ln(n\|h\|)^2}{\sigma^2},$$

The constants are large, but explicit. They can be reduced if stronger conditions than (4.3) are imposed, such as minimal values for n. In typical examples, where $\|h\| \asymp \|\Delta h\| \asymp \sigma^2/n \asymp 1$ we obtain a rate of convergence of $\mathrm{O}(\ln(n)^6 n^{-1/2})$. The correct rate of convergence for the number of empty urns of equiprobable urns is $\mathrm{O}(n^{-1/2})$, which is best possible and was obtained first by Englund (1981). Although our bound does not yield this rate, it is far more general. As a consequence of our result we have the following.

Corollary 4.3. *Sufficient conditions that U as in* (4.2) *is asymptotically normal as $n \to \infty$ are*

(i) $\|\Delta h\|$ *remains bounded,* $\qquad$ (ii) $m\bar{p} = \mathrm{O}(\ln(n))$.

(iii) $\ln(n)^4 n^{2/3}/\operatorname{Var} U \to 0,$ $\qquad$ (iv) $\bar{p} = \mathrm{O}(\ln(n)^{-2})$,

Proof of Theorem 4.2. Let us first state a simple, but key observation, which will be used in the proof:

Fact A. Assume the situation as stated in the theorem and let $K \subset [n]$ be arbitrary. Then, the joint distribution of the balls in the urns of K and K^c is only connected through the total number of balls in these respective subsets. That is, given the number of balls in each urn of K and assuming there are a total of N balls in the urns of K, then the balls in K^c are distributed as if $m - N$ balls were distributed among the urns K^c according to the distribution $\big(p_i/\sum_{k \in K^c} p_k\big)_{i \in K^c}$.

We will use Construction 2F to construct our Stein coupling. First, distribute m balls into n white urns according to $(p_i)_{i \in [n]}$ and denote by ξ_j the number of balls in white urn j. Fix i and note that

$$\mathscr{L}(\xi_i) = \mathrm{Bi}(m, p_i). \qquad (4.4) \quad \{54\}$$

Let us now construct a family of random variables $(\xi'_{i,j})_{j \in [n]}$ such that (3.17) and (3.18) are satisfied.

Assume we have an additional set of n black urns. Independent of all else, let $\xi'_{i,i}$ have distribution (4.4) and put that many balls into black urn i. We proceed with the remaining $\xi'_{i,j}$, $j \neq i$. First, for each $j \neq i$, put ξ_j balls into black urn j. Let $N_1 = |\xi_i - \xi'_{i,i}|$ be the difference of balls in the white and black urn i. If $\xi_i > \xi'_{i,i}$, distribute an additional N_1 balls into the remaining black urns according to the distribution $\big(p_k/(1-p_i)\big)_{k \neq i}$. If $\xi_i < \xi'_{i,i}$, remove instead N_1 balls from the remaining black urns, where the balls are chosen uniformly among all the balls in the black urns except those in black urn i. It is not difficult to see that the construction is in fact symmetric, that is,

$$\mathscr{L}\big((\xi_i)_{i \in [n]}, (\xi'_{i,j})_{j \in [n]}\big) = \mathscr{L}\big((\xi'_{i,j})_{j \in [n]}, (\xi_i)_{i \in [n]}\big).$$

Now, define $U'_i = \sum_j h_i(\xi'_{i,j})$ and $W'_i = (U_i - \mu)/\sigma$. It is clear that (3.17) and (3.18) hold. With I uniformly distributed on $[n]$ and independent of all

and with $\mu_i := \mathbb{E}h_i(\xi_i)$, we hence have that

$$(W, W', G) := (W, W'_I, -n(h_I(\xi_I) - \mu_i)/\sigma)$$

is a Stein coupling.

Fix again i. We now construct another family of random variables $(\xi''_{i,j})_{j \in [n]}$. Denote by $K_i = \{j \neq i : \xi'_{i,j} \neq \xi_j\}$ the set of indices of those urns for which the number of balls in the corresponding black urn and white urn differ, not including urn i, and $\bar{K}_i = K_i \cup \{i\}$. Assume that we have a set of n red urns. First, independently of all else, distribute $N''_2 \sim \text{Bi}(m, \sum_{k \in \bar{K}_i} p_k)$ balls into the red urns from $\bar{K}_i$ according to the distribution $\left(p_j/\sum_{k \in \bar{K}_i} p_k\right)_{j \in \bar{K}_i}$; denote by $N_2 = \sum_{j \in \bar{K}_i} X_j$ the number of balls in the white urns from $\bar{K}_i$, and by $N_3 = |N_2 - N''_2|$ their difference. Now, in very much the same way as we constructed the balls in the black urns from those in the white urns, put ξ_j balls into red urn j for each $j \in \bar{K}^c_i$. If $N_2 > N''_2$ distribute another N_3 balls into the red urns from $\bar{K}^c_i$ according to the distribution $\left(p_j/\sum_{k \in \bar{K}^c_i} p_k\right)_{j \in \bar{K}^c_i}$, and if $N_2 < N''_2$ remove N_3 balls uniformly among all the balls in the red urns from $\bar{K}_i$. Now, given $(\xi_j, \xi'_{i,j})_{j \in \bar{K}_i}$, by construction, the $(\xi''_{j,i})_{j \in [n]}$ always will have the distribution of $(\xi_j)_{j \in [n]}$. Hence, $U''_i := \sum_{j=1}^n h_j(\xi''_{i,j})$ is independent of $U'_i - U = \sum_{j \in \bar{K}_i}\left(h_i(\xi'_{i,j}) - h_i(\xi_i)\right)$ and ξ_i, that is, $W'' = (U''_I - \mu)/\sigma$ is independent of GD.

We are now in a position to apply Theorem 2.5 with $\tilde{D} = D$ and $S = 1$. We clearly have $r_0 = r_1 = r_2 = r_3 = 0$. Note that

$$D = \frac{1}{\sigma} \sum_{j \in \bar{K}_I} \left(h_j(\xi'_{I,j}) - h_j(\xi_j)\right), \qquad D' = \frac{1}{\sigma} \sum_{j \in \bar{K}'_I} \left(h_j(\xi''_{I,j}) - h_j(\xi_j)\right),$$

where $\bar{K}'_i = \bar{K}_i \cup K'_i$ with $K'_i = \{k \in K^c_i : \xi''_k \neq \xi_k\}$.

Let $C > 0$. We have $|\mu_i| \leqslant m\bar{p}\|\Delta h\|$ for every i. Then, with $\alpha := nC\|\Delta h\|/\sigma + nm\bar{p}\|\Delta h\|/\sigma$,

$$\mathbb{P}[|G| > \alpha] \leqslant \mathbb{P}[n|h_I(\xi_I)|/\sigma > \alpha - n\mu_I/\sigma] \leqslant \mathbb{P}[|h_I(\xi_I)| > C\|\Delta h\|]$$
$$\leqslant \mathbb{P}[\xi_I > C] \leqslant \mathbb{P}[\text{Bi}(m, \bar{p}) > C].$$

Furthermore, with $\beta := 2C(C+1)\|\Delta h\|/\sigma$,

$$\mathbb{P}[|D| > \beta] \leqslant \mathbb{P}[\xi'_I \vee \xi_I > C] + \mathbb{P}[|D| > \beta, \xi'_I \vee \xi_I \leqslant C]$$
$$\leqslant \mathbb{P}[\xi'_I \vee \xi_I > C] + \mathbb{P}[\xi'_I \vee \xi_I \leqslant C, \; \exists i \in K_I \text{ s.t. } \xi'_i \vee \xi_i > C]$$
$$+ \mathbb{P}[|D| > \beta, \xi'_i \vee \xi_i \leqslant C \, \forall i \in \bar{K}_I],$$

and, noting that the last expression is zero, this is

$$\leqslant \mathbb{P}[\xi'_I > C] + \mathbb{P}[\xi_I > C] + 2C\mathbb{P}[\text{Bi}(m, \bar{p}/(1-\bar{p})) > C]$$
$$\leqslant 2\mathbb{P}[\text{Bi}(m, \bar{p}) > C] + 2C\mathbb{P}[\text{Bi}(m, 2\bar{p}) > C].$$

Lastly, with $\beta' := 2C(C+1)^2\|\Delta h\|/\sigma$,

$$\mathbb{P}[|D'| > \beta']$$
$$\leqslant \mathbb{P}[\xi'_I \vee \xi_I > C] + \mathbb{P}[|D'| > \beta', \xi'_I \vee \xi_I \leqslant C]$$
$$\leqslant \mathbb{P}[\xi'_I \vee \xi_I > C] + \mathbb{P}\Big[\xi'_I \vee \xi_I \leqslant C, \; \sum_{j \in \bar{K}_I} \xi''_j \vee \sum_{j \in \bar{K}_I} \xi_j > C(C+1)\Big]$$

$$+ \mathbb{P}\big[|D'| > \beta', \xi_I' \vee \xi_I \leqslant C, \sum_{j \in \bar{K}_I} \xi_j'' \vee \sum_{j \in \bar{K}_I} \xi_j \leqslant C(C+1)\big]$$

$$\leqslant \mathbb{P}[\xi_I' \vee \xi_I > C] + \mathbb{P}[\xi_I' \vee \xi_I \leqslant C, \sum_{j \in \bar{K}_I} \xi_j'' \vee \sum_{j \in \bar{K}_I} \xi_j > C(C+1)]$$

$$+ \mathbb{P}\big[\xi_I' \vee \xi_I \leqslant C, \sum_{j \in \bar{K}_I} \xi_j'' \vee \sum_{j \in \bar{K}_I} \xi_j \leqslant C(C+1), \exists i \in \bar{K}_I' \text{ s.t. } \xi_i'' \vee \xi_i > C\big]$$

$$+ \mathbb{P}\big[|D'| > \beta', \xi_I' \vee \xi_I \leqslant C, \xi_i'' \vee \xi_i \leqslant C \, \forall i \in \bar{K}_I'\big],$$

and, noting that the last expression is zero, this is

$$\leqslant 2\mathbb{P}[\mathrm{Bi}(m, \bar{p}) > C] + \mathbb{P}[\mathrm{Bi}(m, (C+1)\bar{p}/(1-\bar{p})) > C(C+1)]$$
$$+ 2C(C+1)\mathbb{P}[\mathrm{Bi}(m, \bar{p}/(1 - C(C+1)\bar{p}) > C]$$
$$\leqslant 2\mathbb{P}[\mathrm{Bi}(m, \bar{p}) > C] + \mathbb{P}[\mathrm{Bi}(m, 2(C+1)\bar{p} > C(C+1)]$$
$$+ 2C(C+1)\mathbb{P}[\mathrm{Bi}(m, 2\bar{p}) > C]$$

if

$$C(C+1)\bar{p} \leqslant 1/2. \qquad (4.5) \quad \{55\}$$

Now, from the Chernoff bound we have that, for every $\varepsilon \geqslant 0$,

$$\mathbb{P}[\mathrm{Bi}(n, p) > (1 + \varepsilon)np] \leqslant \exp\left(-\frac{\varepsilon^2}{2 + \varepsilon} np\right),$$

which implies that

$$\mathbb{P}[\mathrm{Bi}(n, p) > x] \leqslant e^{-x/2} \qquad (4.6) \quad \{56\}$$

as long as $x > 5np$. Choose $C = 4\ln(n\|h\|) - 1$; this implies (4.5) under Condition (4.3). Also under Condition (4.3), (4.6) can be used to obtain

$$\mathbb{P}[|G| > \alpha] \leqslant \mathbb{P}[\mathrm{Bi}(m, \bar{p}) > C] \leqslant \frac{2}{n^2 \|h\|^2},$$

$$\mathbb{P}[|D| > \beta] \leqslant 2\mathbb{P}[\mathrm{Bi}(m, \bar{p}) > C] + 2C\mathbb{P}[\mathrm{Bi}(m, 2\bar{p}) > C]$$
$$\leqslant \frac{4 + 16\ln(n\|h\|)}{n^2 \|h\|^2}$$

$$\mathbb{P}[|D'| > \beta'] \leqslant 2\mathbb{P}[\mathrm{Bi}(m, \bar{p}) > C] + \mathbb{P}[\mathrm{Bi}(m, 2(C+1)\bar{p} > C(C+1)]$$
$$+ 2C(C+1)\mathbb{P}[\mathrm{Bi}(m, 2\bar{p}) > C]$$
$$\leqslant \frac{6 + 64\ln(n\|h\|)^2}{n^2 \|h\|^2}.$$

Now, Theorem 2.5, with α, $\tilde{\beta} := \beta$ and β' as above, and the rough bounds

$$|G| \leqslant \frac{2n\|h\|}{\sigma}, \qquad |D| \leqslant \frac{2n\|h\|}{\sigma}.$$

yields

$$d_{\mathrm{K}}\big(\mathscr{L}(W), \mathrm{N}(0,1)\big)$$

$$\leqslant 12(\alpha\beta + 1)\beta' + 8\alpha\beta^2$$

$$+ \frac{12n^2\|h\|^2}{\sigma^2}\big(\mathbb{P}[|G| > \alpha] + \mathbb{P}[|D| > \beta] + \mathbb{P}[|D'| > \beta']\big)$$

$$\leqslant \frac{48n(C+1)^5\|\Delta h\|^3(C + m\bar{p} + \sigma^2/n)}{\sigma^3} + \frac{32n(C+1)^4\|\Delta h\|^3(C + m\bar{p})}{\sigma^3}$$

$$+ \frac{12\big(12 + 16\ln(n\|h\|) + 64\ln(n\|h\|)^2\big)}{\sigma^2}$$

$$\leqslant \frac{80n(C+1)^5\|\Delta h\|^3(C + m\bar{p} + \sigma^2/n)}{\sigma^3} + \frac{144\big(1 + 11\ln(n\|h\|)^2\big)}{\sigma^2},$$

which, after plugging in C and with some further straightforward manipulations and estimates, yields the final bound. □

4.3. **Neighbourhood statistics of a fixed number of uniformly distributed points.** Consider the space $\mathcal{J} = [0, n^{1/d})^d$ for some integer $d \geqslant 1$ and let $X_1, \ldots, X_n$ be n i.i.d. points uniformly distributed on $\mathcal{J}$. Let ψ be a measurable real-valued functional defined on all pairs $(x, \mathcal{X})$ where $x \in \mathcal{J}$ and $\mathcal{X} \subset \mathcal{J}$ is a finite subset. Such statistics have been investigated in many places for specific choices of ψ. If the number of points is Poisson distributed, approaches using local dependence can be successfully applied. But if the number of points is fixed, besides the local dependence also global weak dependence has to be taken into account.

Assume that ψ is translation invariant, i.e. $\psi(x + y, \mathcal{X} + y) = \psi(x, \mathcal{X})$, where we assume the torus convention for translations. Assume that ψ has influence radius ρ, i.e. for each $x \in \mathcal{J}$ and each $\mathcal{X} \subset \mathcal{J}$ we have

$$\psi(x, \mathcal{X}) = \psi(x, \mathcal{X} \cap B_\rho(x)),$$

where $B_\rho(x)$ is the closed sphere of radius ρ and center x under the toroidal Euclidean metric. To avoid self-overlaps, assume $\rho < \frac{1}{2}n^{1/d}$. Let $\mathcal{X} := \{X_1, \ldots, X_n\}$, define $U = \sum_{x \in \mathcal{X}} \psi(x, \mathcal{X})$, $\sigma^2 = \mathrm{Var}\, U$ and $W = U/\sigma$, where we assume that $\mathbb{E}\psi(X_1, \mathcal{X}) = 0$.

Theorem 4.4. *Assume W is defined as above and assume also that $\rho \geqslant \pi^{-1/2}\Gamma(1 + d/2)^{1/d}$. Then there is a universal constant C_d only depending on d such that*

$$d_{\mathrm{W}}\big(\mathscr{L}(W), \mathrm{N}(0,1)\big) \leqslant \frac{C_d\|\psi\|^3\rho^{6d}n}{\sigma^3}.$$

In the case where d, ρ and $\|\psi\|$ remain fixed as $n \to \infty$ we have from Penrose and Yukich (2001) that $\sigma^2 \asymp n$ as long as $\mathrm{Var}(\psi(X_1, \mathcal{X})) > 0$; in that case, Theorem 4.4 gives the best possible order $n^{-1/2}$ for the Wasserstein metric.

After constructing an extended Stein coupling, the proof of the theorem essentially amounts to bounding the third moments of some mixed binomial distributions. Constants could be easily extracted from the proof but with the rough bound given here they are too large to be of practical use.

Proof of Theorem 4.4. We first make a simple, but key observation, which will be used in the proof:

Fact B. Let $\mathcal{Y}$ be a fixed number of points distributed uniformly and independently of each other on an open subset $\mathcal{J}' \subset \mathcal{J}$. For any other open subset $U \subset \mathcal{J}$, the joint distribution of points of $\mathcal{Y}$ on U and U^c is only connected through the number of points in these two respective subsets via the equation $|U \cap \mathcal{Y}| + |U^c \cap \mathcal{Y}| = |\mathcal{Y}|$. That is, given the total number and the number of points in one of the subsets, U, say, possibly along with their locations on $U \cap \mathcal{J}'$, the remaining points in the subset U^c are then distributed uniformly and independently of each other on $U^c \cap \mathcal{J}'$.

We first construct a Stein coupling according to Construction 2A. Note that W_i' constructed below will not have the same distribution as W. Let to this end $I = i$ and $X_i = x_i$ be given. Let

$$N_{i,1} = \big(\mathcal{X} \cap B_\rho(x_i)\big) \setminus \{x_i\}$$

be the points within radius ρ of x_i, excluding x_i. Let $N_{i,2}$ be an $|N_{i,1}|$ number of points uniformly distributed on $B_\rho^c(x_i)$. Define the new set

$$\mathcal{X}_i' = \big(\mathcal{X} \cap B_\rho^c(x_i)\big) \cup N_{i,2};$$

that is, remove x_i from $\mathcal{X}$ and replace the neighbouring points of x_i by the new points $N_{i,2}$. Note that $|\mathcal{X}_i'| = n - 1$. Now, let

$$U_i' = \sum_{x \in \mathcal{X}_i'} \psi(x, \mathcal{X}_i').$$

As $\psi(x_i, \mathcal{X}) = \psi(x_i, \mathcal{X} \cap B_\rho(x_i))$ and using Fact B, we have that S_i' is independent of $\psi(x_i, \mathcal{X})$.

Randomising now over I and X_I, define $U' = U_I'$, $W' = U_I'/\sigma$ and $G = -n\psi(X_I, \mathcal{X})$. Then, from Construction 2A, we have that (W, W', G) is a Stein coupling. Indeed, using the above mentioned independence,

$$\mathbb{E}\big\{\psi(X_i, \mathcal{X}) \,\big|\, U_i'\big\} = 0,$$

and hence (3.13) is satisfied.

We now construct an extension W'' of the basic Stein coupling, which will be independent of G and $U' - U$. Let again $I = i$ and $X_i = x_i$ be given. First, we define some further sets. Define

$$M_{i,1} = B_{2\rho}(x_i) \cup \bigcup_{x \in N_{i,2}} B_\rho(x)$$

and denote by $N_{i,3} = \mathcal{X} \cap (M_{i,1} \setminus B_\rho(x_i))$ all the points of $\mathcal{X}$ which remained fixed during the perturbation, but whose values of ψ were potentially affected, either because of points being removed or points being added in their neighbourhood. We can now write

$$\Delta_i := U_i' - U = \sum_{x \in N_{i,2}} \psi(x, \mathcal{X}_i') + \sum_{x \in N_{i,3}} \big(\psi(x, \mathcal{X}_i') - \psi(x, \mathcal{X})\big)$$
$$- \sum_{x \in N_{i,1}} \psi(x, \mathcal{X}) - \psi(x_i, \mathcal{X}). \tag{4.7} \quad \{57\}$$

Let

$$M_{i,2} = B_{3\rho}(x_i) \cup \bigcup_{x \in N_{i,2}} B_{2\rho}(x).$$

Clearly, the values of Δ_i and $\psi(x_i, \mathcal{X})$ are determined by the points $\mathcal{X} \cap M_{i,2}$ and $\mathcal{X}' \cap M_{i,2}$ only. Given the points $\mathcal{X} \cap M_{i,2}$ and $\mathcal{X}' \cap M_{i,2}$, we have from Fact A that the remaining $n - |\mathcal{X} \cap M_{i,2}|$ points of $\mathcal{X}$ are distributed uniformly and independently of each other on $M_{i,2}^c$. Hence we can use them for our construction of W'' as follows. First, let $N_{i,4}$ be a $\mathrm{Bi}(n, \mathrm{Vol}(M_{i,2})/n)$ number of points distributed independently and uniformly on $M_{i,2}$. Then, let $N_{i,5}$ be the remaining $n - |N_{i,4}|$ points on $M_{i,2}^c$ distributed as follows. Let $P(k)$ be k random points distributed uniformly on $M_{i,2}^c$ if $k > 0$, and let, for $k < 0$, $P(k)$ be $|k|$ randomly chosen points from $\mathcal{X} \cap M_{i,2}^c$ (without replacement); set $P(0) = \emptyset$. Let $P := P(|\mathcal{X} \cap M_{i,2}| - |N_{i,4}|)$ and define

$$N_{i,5} = \begin{cases} (\mathcal{X} \cap M_{i,2}^c) \setminus P & \text{if } |N_{i,4}| > |\mathcal{X} \cap M_{i,2}|, \\ (\mathcal{X} \cap M_{i,2}^c) \cup P & \text{if } |N_{i,4}| \leqslant |\mathcal{X} \cap M_{i,2}|; \end{cases}$$

that is, we let $N_{i,5}$ be the points $\mathcal{X} \cap M_{i,2}^c$ and add or remove as many points (that is the points P) as needed so that $|N_{i,4}| + |N_{i,5}| = n$. Now, setting $\mathcal{X}_i'' = N_{i,4} \cup N_{i,5}$, we clearly have that $U_i'' = \sum_{x \in \mathcal{X}_i''} \psi(x, \mathcal{X}_i'')$ is independent of $(\psi(x_i, \mathcal{X}), \Delta)$ as the conditional distribution of $\mathcal{X}_i''$ given $(\mathcal{X} \cap M_{i,2}, \mathcal{X}_i' \cap M_{i,2})$ always equals to $\mathscr{L}(\mathcal{X})$ by construction and $(\mathcal{X} \cap M_{i,2}, \mathcal{X}_i' \cap M_{i,2})$ determines $(\psi(x_i, \mathcal{X}), \Delta_i)$ as mentioned before. Randomizing over I and X_I, set $U'' = U_I''$ and $W'' = U''/\sigma$. Thus, (W, W', G, W'') satisfies the assumptions of Corollary 2.4 when setting $\tilde{D} = D$. It remains to find the corresponding quantities A and B.

Define the sets

$$M_{i,3} = B_{4\rho}(x_i) \cup \bigcup_{x \in N_{i,2}} B_{3\rho}(x),$$

$$M_{i,4} = \bigcup_{x \in P} B_\rho(x),$$

and let

$$\begin{aligned} Y_1 &:= |\mathcal{X} \cap (B_\rho(x_i) \setminus \{x_i\})| = |N_{i,1}| = |N_{i,2}| \\ Y_2 &:= |\mathcal{X} \cap (M_{i,2} \setminus B_\rho(x_i))| = |\mathcal{X} \cap M_{i,2}| - 1 - Y_1, \\ Y_3 &:= |\mathcal{X} \cap (M_{i,3} \setminus M_{i,2})| = |\mathcal{X} \cap M_{i,3}| - 1 - Y_1 - Y_2, \\ Y_4 &:= |N_{i,4}| = |\mathcal{X}_i'' \cap M_{i,2}|. \end{aligned}$$

Let $\kappa_\rho := \mathrm{Vol}(B_\rho(0))$. Then the following statements are straightforward:

(i) $\mathscr{L}(Y_1) = \mathrm{Bi}(n-1, \kappa_\rho/n)$,

(ii) $\mathscr{L}(Y_2 + Y_3 | N_{i,2}) = \mathrm{Bi}\left(n - 1 - Y_1, \frac{\mathrm{Vol}(M_{i,3} \setminus B_\rho(x_i))}{(n - \kappa_\rho)}\right)$,

(iii) $\mathscr{L}(Y_4 | N_{i,2}) = \mathrm{Bi}\left(n, \mathrm{Vol}(M_{i,2})/n\right)$,

(iv) $Y_4 \perp\!\!\!\perp (Y_2, Y_3)$ given $N_{i,2}$,

(v) $|P| \leqslant |\mathcal{X} \cap M_{i,2}| + |\mathcal{X}_i'' \cap M_{i,2}| = 1 + Y_1 + Y_2 + Y_4$,

(vi) $|\mathcal{X}_i'' \cap M_{i,3}| \leqslant Y_4 + Y_3 + |P| \leqslant 1 + Y_1 + Y_2 + Y_3 + 2Y_4$,

(vii) $|(\mathcal{X}_i'' \setminus P) \cap M_{i,4}| \leqslant |\mathcal{X}_i'' \cap M_{i,3}| + Y_5$ and

$\qquad |(\mathcal{X} \setminus P) \cap M_{i,4}| \leqslant |\mathcal{X}_i \cap M_{i,3}| + Y_5,$

$\qquad$ for some Y_5 with $\mathscr{L}(Y_5|P, N_{i,2}, N_{i,4}) \sim \mathrm{Bi}\left(n - Y_1 - 1, \frac{\mathrm{Vol}(M_{i,4})}{n - \mathrm{Vol}(M_{i,3})}\right),$

The bound in (vi) follows from the fact that the difference between the points $\mathcal{X} \cap (M_{i,3} \setminus M_{i,2})$ and $\mathcal{X}'' \cap (M_{i,3} \setminus M_{i,2})$ is at most the points from P. The somewhat rough bound (vii) is due to the fact that the neighbourhoods of the points in P may overlap with $M_{i,3}$.

Let now C be a constant that may depend only on the dimension d but can change from formula to formula. From representation (4.7), we easily obtain

$$|D_i| \leqslant \|\psi\|\big(|N_{i,2}| + 2|N_{i,3}| + |N_{i,1}| + 1\big)/\sigma \leqslant C\|\psi\|(Y_1 + Y_2 + 1)/\sigma.$$

Now, we can write $\Delta_i' := U_i'' - U$ as

$$\Delta_i' = \sum_{x \in \mathcal{X}_i'' \cap M_{i,3}} \varphi(x, \mathcal{X}_i'') - \sum_{x \in \mathcal{X} \cap M_{i,3}} \varphi(x, \mathcal{X})$$
$$+ \sum_{x \in \mathcal{X}_i'' \cap (M_{i,4} \setminus M_{i,3})} \varphi(x, \mathcal{X}_i'') - \sum_{x \in \mathcal{X} \cap (M_{i,4} \setminus M_{i,3})} \varphi(x, \mathcal{X}),$$

hence

$$|D_i'| \leqslant \|\psi\|\big(|\mathcal{X}_i'' \cap M_{i,3}| + |\mathcal{X} \cap M_{i,3}| + |\mathcal{X}_i'' \cap M_{i,4}| + |\mathcal{X} \cap M_{i,4}|\big)/\sigma$$
$$\leqslant C\|\psi\|Z/\sigma,$$

where $Z := 1 + Y_1 + Y_2 + Y_3 + Y_4 + Y_5$.

To estimate the third absolute moment of Z note first that if $Y \sim \mathrm{Bi}(m, p)$ with $mp \geqslant 1$ we have

$$\mathbb{E}Y^3 \leqslant 5m^3 p^3.$$

Note that the assumption $\rho \geqslant \pi^{-1/2}\Gamma(1+d/2)^{1/d}$ implies that $\mathbb{E}Y_1 = \kappa_\rho \geqslant 1$ because $\kappa_\rho = \rho^d \kappa_1 = \rho^d \pi^{d/2}/\Gamma(1+d/2)$.

As the cubic function is convex on the non-negative half line, we have for any non-negative numbers $a_1, \ldots, a_m$ that

$$(a_1 + \cdots + a_m)^3 \leqslant m^2(a_1^3 + \cdots + a_m^3).$$

From (i) we immediately obtain

$$\mathbb{E}Y_1^3 \leqslant 5\kappa_\rho^3. \qquad (4.8) \quad \{58\}$$

From (ii)–(iv) we have that, given Y_1, $Y_2 + Y_3 + Y_4$ is stochastically dominated by $\mathrm{Bi}(2n, 2\,\mathrm{Vol}(M_{i,3})/n)$ which is further stochastically dominated by $\mathrm{Bi}(2n, (2\kappa_{4\rho} + 2\kappa_{3\rho}Y_1)/n)$, where we define $\mathrm{Bi}(m, p) := \mathrm{Bi}(m, 1)$ if $p > 1$; hence, using this and (4.8),

$$\mathbb{E}(Y_2 + Y_3 + Y_4)^3 \leqslant C\mathbb{E}(\kappa_{4\rho} + \kappa_{3\rho}Y_1)^3 \leqslant C(\kappa_\rho^3 + \kappa_\rho^6). \qquad (4.9) \quad \{59\}$$

Note now that, given P, $N_{i,2}$ and $N_{i,4}$, we have from (vii) that $\mathscr{L}(Y_5)$ is stochastically dominated by $\mathrm{Bi}(n - 1 - Y_1, \kappa_\rho|P|/(n - \kappa_{4\rho} - \kappa_{3\rho}Y_1)_+)$ which,

by (v), can be dominated by $\mathrm{Bi}(n - 1 - Y_1, \kappa_\rho(1 + Y_1 + Y_2 + Y_4)/(n - \kappa_{4\rho} - \kappa_{3\rho}Y_1)_+)$; hence, using this, (4.8) and (4.9), we obtain

$$\mathbb{E}Y_5^3 \leqslant C\mathbb{E}\left\{(n - 1 - Y_1)\left(1 \wedge \frac{\kappa_\rho^3(1 + Y_1 + Y_2 + Y_4)}{(n - \kappa_{4\rho} - \kappa_{3\rho}Y_1)_+}\right)\right\}^3$$
$$\leqslant n\mathbb{P}[Y_1 > cn] + C\kappa_\rho^3\mathbb{E}(1 + Y_1 + Y_2 + Y_4)^3$$

for n large enough and c such that $n - \kappa_{4\rho} - cn\kappa_{3\rho} \geqslant n/2$, e.g. $c = 1/(4\kappa_{3\rho})$ and $n \geqslant 4\kappa_{4\rho}$. From the Chernoff-Hoeffding inequality we obtain that $\mathbb{P}[Y_1 \geqslant cn] \leqslant 2^{-cn}$ for cn large enough. Hence, $n\mathbb{P}[Y_1 \geqslant nc] \leqslant C \leqslant C\kappa_\rho^3$ and

$$\mathbb{E}Y_5^3 \leqslant C\kappa_\rho^3\left(1 + \mathbb{E}Y_1^3 + \mathbb{E}(Y_2 + Y_3 + Y_4)^3\right)$$
$$\leqslant C(\kappa_\rho^3 + \kappa_\rho^6 + \kappa_\rho^9). \tag{4.10} \quad \{60\}$$

Putting (4.8)–(4.10) together we obtain

$$\mathbb{E}Z^3 \leqslant C(1 + \mathbb{E}Y_1^3 + \mathbb{E}(Y_2 + Y_3 + Y_4)^3 + \mathbb{E}Y_5^3)$$
$$\leqslant C(1 + \kappa_\rho^3 + \kappa_\rho^6 + \kappa_\rho^9) \leqslant C\kappa_\rho^9,$$

as $\kappa_\rho \geqslant 1$. Hence

$$\mathbb{E}|D'|^3 = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}|D_i'|^3 \leqslant C\|\psi\|^3\kappa_\rho^9/\sigma^3 =: A^3.$$

Furthermore, we have

$$\mathbb{E}|G|^3 \leqslant n^3\|\psi\|^3/\sigma^3 =: B^3. \tag{4.11} \quad \{61\}$$

Now, Corollary 2.4 yields

$$d_{\mathrm{W}}\big(\mathscr{L}(W), \mathrm{N}(0,1)\big) \leqslant 5A^2B \leqslant Cn\|\psi\|^3\kappa_\rho^6/\sigma^3.$$

By recalling that $\kappa_\rho = \rho^d\kappa_1$ we obtain the final bound. $\qquad\square$

4.4. Susceptibility and related statistics in the sub-critical Erdős-Rényi random graph.

Let H be a graph (here, we use the letter 'H' for graphs as the letter 'G' is used in the context of Stein couplings). Then, *susceptibility* $\chi(H)$ is defined to be the expected size of the component containing a uniformly chosen random vertex. That is, if $C_i \subset H$, $i = 1, \ldots, K$ are the K maximal subgraphs of the graph H,

$$\chi(H) = \sum_{i=1}^{K}\frac{|C_i|}{n}|C_i|, \tag{4.12} \quad \{62\}$$

where $|C_i|$ denotes the number of vertices in subgraph C_i. Using a different normalisation in (4.12) we can also write

$$\sum_{i=1}^{K}\frac{|C_i|}{n}\frac{|C_i|}{n} = \frac{1}{n}\chi(H), \tag{4.13} \quad \{63\}$$

which, hence, is the probability that, for a given graph H, two randomly chosen vertices are in the same component, or, equivalently, connected through a path. Note that, as H is random, the probability of being connected is therefore itself a random quantity.

Using a martingale CLT it was proved by Janson and Luczak (2008) that, if H is a sub-critical Erdős-Rényi graph, $\chi(H)$ is asymptotically normal. However, it is well known that standard martingale CLTs will often not yield optimal rates of convergence with respect to the Kolmogorov metric (see e.g. Bolthausen (1982)). As often the actual dependence structure under consideration is much weaker than in a worst-case scenario, Rinott and Rotar' (1998) provide bounds that incorporate expressions to measure the dependency and which yield optimal bounds in many settings, but the involved quantities are typically hard to calculate. Let us derive here some bounds without using martingales and which are optimal up to some additional logarithmic factors.

Let us consider here more general statistics that depend only on the properties of the components of two randomly chosen vertices. For a vertex $i \in V(H)$ denote by $C(H, i) \subset H$ the component (i.e. the maximal subgraph) containing i. Let $h = h(H, i, j)$ be a function that is determined by i and j and the two components that contain i and j, respectively, that is,

$$h(H, i, j) = h(H \cap (C(H, i) \cup C(H, j)), i, j), \qquad \text{for all } H, i \text{ and } j, \quad (4.14) \quad \{64\}$$

and which is symmetric in the sense that

$$h(H, i, j) = h(H, j, i), \qquad \text{for all } H, i \text{ and } j. \qquad (4.15) \quad \{65\}$$

Then define

$$U(H) := \sum_{i \in V(H)} \sum_{j \in V(H)} h(H, i, j). \qquad (4.16) \quad \{66\}$$

We can recover susceptibility (4.12) from (4.16) (up to a normalising constant) simply by choosing

$$h(H, i, j) = \mathrm{I}[i \text{ and } j \text{ are in the same component}]. \qquad (4.17) \quad \{67\}$$

As there is little hope for a normal limit for general h we consider functions that are non-zero only if the two random vertices are in the same component, that is,

$$C(H, i) \neq C(H, j) \quad \Longrightarrow \quad h(H, i, j) = 0. \qquad (4.18) \quad \{68\}$$

Hence, under (4.18), we can write (4.16) also as

$$U(H) = \sum_{i \in V(H)} \sum_{j \in C(H, i)} h(H, i, j). \qquad (4.19) \quad \{69\}$$

Dependence on just one vertex and its component is, of course, also covered as a special case $h(H, i, i) = h_0(H, i)$ and $h(H, i, j) = 0$ if $i \neq j$. For example $h_0(H, i) = \mathrm{I}[|C(H, i)| = 1]$ yields the total number of singletons, or, more generally, choose $h_0(H, i) = \mathrm{I}[C(H, i) > m_0]$ to obtain the number of vertices that are in a component of size larger than m_0. We can also recover (4.12) by setting $h_0(H, i) = |C(H, i)|^2/n$, however this function is not Lipschitz in the size of the component. Other quantities of interest may be obtained, such as

$h(H, i, j) = \mathrm{I}[i \text{ and } j \text{ are connected, but not further than } m_0 \text{ apart}],$

$h(H, i, j) = \mathrm{I}[i \text{ and } j \text{ are in the same cycle}],$

$h(H, i, j) = \mathrm{I}[i \text{ and } j \text{ are connected}]/|C(H, i)|.$

Note that by summing uniformly over all the vertices, the components are size-biased. However, we can 'de-bias' as follows: for a given $h_0(H,i)$ of interest, $h_0^u(H,i) = h_0(H,i)/|C(H,i)|$ gives the corresponding unbiased result, where now summing is done uniformly among the components. As $|C(H,i)| \geqslant 1$, we have $\|\Delta h_0^u\| \leqslant \|\Delta h_0\|$. However, averaging is only possible with respect to the expected number of components, that is, $h_0^u(H,i) = \frac{h_0(H,i)}{|C(H,i)|\mathbb{E}K}$, not the actual number, as the function $\frac{h_0(H,i)}{|C(H,i)|K}$ does not satisfy (4.19).

We have the following theorem.

Theorem 4.5. *Let U be as in (4.16) for some function h satisfying (4.14), (4.15) and (4.18). Let $W = (U(H) - \mu)/\sigma$, where $\mu = \mathbb{E}U(H)$ and $\sigma^2 = \operatorname{Var} U(H)$. Let*

$$\|h\| = \sup_H \sup_{i,j \in [n]} |h(H,i,j)| \geqslant 1$$

and, with $H^{k,l}$ denoting the graph where an additional edge is added between k and l if $k \neq l$,

$$\|\Delta h\| = \sup_{i,j,k,l \in [n]} \sup_H \left| h(H^{k,l}, i, j) - h(H, i, j) \right| \geqslant 1.$$

Assume that $0 \leqslant h(H,i,j) \leqslant 2\|\Delta h\|$ whenever H is such that the components which contain i and j, respectively, are singletons. Let H be an Erdős-Rényi random graph with n vertices and edge probabilities $\lambda/n < 1/n$. Then, there is a universal constant $K > 0$ such that

$$d_K\big(\mathscr{L}(W), \mathrm{N}(0,1)\big)$$
$$\leqslant K\left(\frac{n \log(n\|h\|)^{11}\|\Delta h\|^3(1 + 1/a^2 + \sigma^2/n)}{\lambda a^{11}\sigma^3} + \frac{e^a \ln(n\|h\|)^2}{\lambda a^2 \sigma^2} \right)$$

whenever

$$a := \lambda - 1 - \log \lambda \leqslant 4\ln(n\|h\|). \tag{4.20} \quad \{70\}$$

Let now $\lambda < 1$ be fixed and let h be as in (4.17) to obtain susceptibility (up a normalising constant). Clearly, $\|h\| = 1$ and $\|\Delta h\| = 1$. From Janson and Luczak (2008) we have that $\operatorname{Var} U(H) \sim 2\lambda n(1-\lambda)^{-5}$, hence Theorem 4.5 yields a Kolmogorov bound of order $\log(n)^{11}/\sqrt{n}$.

Proof of Theorem 4.5. We will make use of Construction 2A to obtain a Stein coupling. We consider throughout the vertex set $[n]$. Let $C_V(H,i)$ denote the vertices of $C(H,i)$. If $e = \{k,l\} \subset V(H) = [n]$ is a potential edge in a graph H on $[n]$ and $K \subset [n]$, write $e \sim K$ if $k \in K$ or $l \in K$ (or both). Write $e \not\sim K$ if $k \notin K$ and $l \notin K$.

Let now $\xi = (\xi_{\{i,j\}})$, where $i,j \in [n]$ and $i \neq j$, be an i.i.d. family of $\mathrm{Be}(\lambda/n)$ distributed random variables. Let H be the graph on the vertex set $[n]$ with edge set ξ, that is, e is an edge in H iff $\xi_e = 1$. Let H^* be an independent copy of H. Note that, given $C(H,i)$, we have that $(\xi_e)_{e \not\sim C_V(H,i)}$ is a family of i.i.d. $\mathrm{Be}(\lambda/n)$ random variables. Define the random graph H_i' through

$$e \in E(H_i') \quad :\Longleftrightarrow \quad \begin{array}{l} e \in E(H^*) \text{ and } e \sim C_V(H,i), \text{ or} \\ e \in E(H) \text{ and } e \not\sim C_V(H,i). \end{array}$$

As H_i', given $C(H, i)$, is always an unconditional Erdős-Rényi random graph by construction, we have that $C(H, i)$ is independent of the graph H_i'.

Now, let I be uniformly and independently distributed on $[n]$. Set $H' = H_I'$, $U = U(H)$ and $U' = U(H')$. Furthermore, $W = (U - \mu)/\sigma$, $W' = (U' - \mu)/\sigma$ and

$$G = -\frac{n}{\sigma}\left(\sum_{j \in C(H,I)} h(H, I, j) - \mu_I\right),$$

where $\mu_i := \mathbb{E}\sum_{j \in C(H,i)} h(H, i, j)$. Then, it is easy to see that (3.13) is satisfied and hence we have a Stein coupling.

Let now

$$V_i := \bigcup_{k \in C_V(H,i)} C_V(H_i', k)$$

be all those vertices of the graph H whose components are affected by changing from H to H_i'; that is, the vertices $C_V(H, i)$ themselves on one hand and, on the other hand, the vertices $[n] \setminus C_V(H, i)$ that got connected in H_i' to at least one of the vertices $C_V(H, i)$.

Then we can write

$$\Delta_i := U(H_i') - U(H) = \sum_{k \in V_i}\sum_{l \in C_V(H,k)}\big(h(H_i', k, l) - h(H, k, l)\big).$$

Note that the subgraphs $H \cap V_i$ and $H_i' \cap V_i$ determine the value of Δ_i. Hence, similarly as before, let H^{**} be an independent copy of H and define the graph H_i'' by

$$e \in E(H_i'') \quad :\iff \quad \begin{matrix} e \in E(H^{**}) \text{ and } e \sim V_i, \text{ or} \\ e \in E(H) \text{ and } e \nsim V_i. \end{matrix}$$

With the same argument as before, H_i'' is independent of $H \cap V_i$ and $H_i' \cap V_i$. Define

$$V_i' = \bigcup_{k \in V_i} C_V(H_i'', k)$$

to be all those vertices of the graph H whose components are affected by changing from H to H_i''. We can write

$$\Delta_i' := U(H_i'') - U(H) = \sum_{k \in V_i'}\sum_{l \in C_V(H,k)}\big(h(H_i'', k, l) - h(H, k, l)\big).$$

We are now in a position to apply Theorem 2.5. As (W, W', G) is a Stein coupling we have $r_0 = 0$. Setting $\breve{D} = D$ and $S = 1$ we have $r_1 = r_2 = 0$, and by construction of H_i'', that $r_3 = 0$.

Let $C > 0$ to be chosen later. From Durrett (2007, p. 38) we have that

$$\mathbb{P}(|C(H, 1)| \geqslant k) \leqslant \frac{e^{-ak}}{\lambda}. \tag{4.21} \quad \{71\}$$

Hence, we obtain the simple estimate

$$\mathbb{E}|C(H, 1)|^2 = \int_0^\infty \mathbb{P}[|C(H, 1)| > \sqrt{x}]dx \leqslant \frac{2}{\lambda a^2}.$$

Note also that, if $j \in C_V(H, i)$,

$$|h(H, i, j)| \leqslant \|\Delta h\||C(H, i)|.$$

This yields $|\mu_i| \leqslant \|\Delta h\| \mathbb{E}|C(H,i)|^2 \leqslant \|\Delta h\| \frac{2}{\lambda a^2}$. Furthermore, if $j \in V_i \setminus C_V(H,i)$, then

$$|C(H,j)| \leqslant |C(H'_i,j)|.$$

With $\alpha := \frac{n}{\sigma}\|\Delta h\|(C^2 + \frac{2}{\lambda a^2})$ we have

$$\mathbb{P}[|G| > \alpha] \leqslant \mathbb{P}[|C(H,I)| > C] + \mathbb{P}[|G| > \alpha, |C(H,I)| \leqslant C]$$

and, noticing that the last term is zero, this is

$$= \mathbb{P}[|C(H,1)| > C].$$

Furthermore, with $\beta := 2C^4\|\Delta h\|/\sigma$,

$$
\begin{aligned}
&\mathbb{P}[|D| > \beta]\\
&\leqslant \mathbb{P}[|C(H,I)| > C] + \mathbb{P}[|D| > \beta, |C(H,I)| \leqslant C]\\
&\leqslant \mathbb{P}[|C(H,1)| > C]\\
&\qquad + \mathbb{P}[|C(H,I)| \leqslant C, \exists j \in C_V(H,I) \text{ s.t. } |C(H'_I,j)| > C]\\
&\qquad + \mathbb{P}[|D| > \beta, |C(H,I)| \leqslant C, \ |C(H'_I,j)| \leqslant C \text{ for all } j \in C_V(H,I)]
\end{aligned}
$$

and, noticing that the last term is zero, this is

$$
\begin{aligned}
&\leqslant \mathbb{P}[|C(H,1)| > C] + C\mathbb{P}[|C(H,1)| > C]\\
&= (C+1)\mathbb{P}[|C(H,I)| > C]
\end{aligned}
$$

Finally, with $\beta' = 2C^5\|\Delta h\|/\sigma$,

$$
\begin{aligned}
&\mathbb{P}[|D'| > \beta']\\
&\leqslant \mathbb{P}[|C(H,I)| > C] + \mathbb{P}[|D'| > \beta', |C(H,I)| \leqslant C]\\
&\leqslant \mathbb{P}[|C(H,1)| > C]\\
&\qquad + \mathbb{P}[|C(H,I)| \leqslant C, \exists j \in C_V(H,I) \text{ s.t. } |C(H'_I,j)| > C]\\
&\qquad + \mathbb{P}[|D'| > \beta', |C(H,I)| \leqslant C, \ |C(H'_I,j)| \leqslant C \ \forall j \in C_V(H,I)]\\
&\leqslant \mathbb{P}[|C(H,1)| > C] + C\mathbb{P}[|C(H,1)| > C]\\
&\qquad + \mathbb{P}\big[|C(H,I)| \leqslant C, \ |C(H'_I,j)| \leqslant C \ \forall j \in C_V(H,I),\\
&\qquad\qquad\qquad\qquad\qquad\qquad \exists j \in V_I \text{ s.t. } |C(H''_I,j)| > C\big]\\
&\qquad + \mathbb{P}\big[|D'| > \beta', |C(H,I)| \leqslant C, \ |C(H'_I,j)| \leqslant C \ \forall j \in C_V(H,I),\\
&\qquad\qquad\qquad\qquad\qquad\qquad |C(H''_I,j)| \leqslant C \ \forall j \in V_I\big]
\end{aligned}
$$

and, noticing that the last term is zero, this is

$$
\begin{aligned}
&\leqslant \mathbb{P}[|C(H,1)| > C] + C\mathbb{P}[|C(H,1)| > C] + C^2\mathbb{P}[|C(H,1)| > C]\\
&\leqslant (C+1)^2\mathbb{P}[|C(H,1)| > C].
\end{aligned}
$$

We also have the rough bounds

$$|G|, |D| \leqslant \frac{2n^2}{\sigma}\|h\|,$$

Hence, choosing $C = 8\ln(n\|h\|)/a - 1$ we have that $C \geqslant 1$ from (4.20), and using (4.21) we have

$$\mathbb{P}[|G| > \alpha] \leqslant \frac{e^{-aC}}{\lambda} \leqslant \frac{e^a}{\lambda n^4 \|h\|^2},$$

$$\mathbb{P}[|D| > \beta] \leqslant (C+1)\frac{e^{-aC}}{\lambda} \leqslant \frac{4e^a \ln(n\|h\|)}{a\lambda n^4 \|h\|^2},$$

$$\mathbb{P}[|D'| > \beta'] \leqslant (C+1)^2 \frac{e^{-aC}}{\lambda} \leqslant \frac{16 e^a \ln(n\|h\|)^2}{a^2 \lambda n^4 \|h\|^2}.$$

From Theorem 2.5 we hence have

$$d_{\mathrm{K}}(\mathscr{L}(W), \mathrm{N}(0,1))$$

$$\leqslant 12(\alpha\beta + 1)\beta' + 8\alpha\beta^2$$

$$+ \frac{4n^4\|h\|^2}{\sigma^2}\big(\mathbb{P}[|G| > \alpha] + \mathbb{P}[|D| > \beta] + \mathbb{P}[|D'| > \beta']\big)$$

$$\leqslant 12\left(\frac{n\|\Delta h\|}{\sigma}\left(C^2 + \frac{2}{\lambda a^2}\right)\frac{2C^4\|\Delta h\|}{\sigma} + 1\right)\frac{2C^5\|\Delta h\|}{\sigma}$$

$$+ 8\frac{n\|\Delta h\|}{\sigma}\left(C^2 + \frac{2}{\lambda a^2}\right)\frac{4C^8\|\Delta h\|^2}{\sigma^2}$$

$$+ \frac{4n^4\|h\|^2}{\sigma^2}\left(\frac{e^2}{\lambda n^4\|h\|^2} + \frac{4e^a\ln(n\|h\|)}{a\lambda n^4\|h\|^2} + \frac{16e^a\ln(n\|h\|)^2}{a^2\lambda n^4\|h\|^2}\right)$$

$$\leqslant 24\left(\left(C^2 + \frac{2}{\lambda a^2}\right)C^4 + \frac{\sigma^2}{n}\right)\frac{nC^5\|\Delta h\|^3}{\sigma^3}$$

$$+ 32\left(C^2 + \frac{2}{\lambda a^2}\right)\frac{C^8 n\|\Delta h\|^3}{\sigma^3} + \frac{84e^a\ln(n\|h\|)^2}{a^2\lambda\sigma^2}$$

$$\leqslant 32\left(2 + \frac{4}{a^2\lambda} + \frac{\sigma^2}{n}\right)\frac{nC^{11}\|\Delta h\|^3}{\sigma^3} + \frac{84e^a\ln(n\|h\|)^2}{\lambda a^2\sigma^2}.$$

$\square$

Remark 4.2. Note that (H, H'_i) does not form an exchangeable pair, although the marginal distributions are the same. To see this, denote by G_c the complete graph and by G_0 the empty graph on the vertices $[n]$, where we assume $n > 2$. Now, given $H = G_c$, H'_i is just an independent realisation of the Erdős-Rényi random graph, hence

$$\mathbb{P}[H'_i = G_0 | H = G_c] = \mathbb{P}[H'_i = G_0] = (1-p)^{\binom{n}{2}}.$$

On the other hand,

$$\mathbb{P}[H = G_0 | H'_i = G_c] = \frac{\mathbb{P}[H'_i = G_c | H = G_0]\mathbb{P}[H = G_0]}{\mathbb{P}[H'_i = G_c]} = 0$$

since it is not possible that H'_i is a complete graph if H is empty.

5. ZERO BIAS TRANSFORMATION

Assume that $\mathbb{E}W = 0$ and $\mathrm{Var}\,W = 1$. It was proved in Goldstein and Reinert (1997) that there exists a unique distribution $\mathscr{L}(W^z)$ such that, for all

smooth f we have

$$\mathbb{E}\{Wf(W)\} = \mathbb{E}f'(W^z) \qquad (5.1) \quad \{72\}$$

Furthermore, $\mathscr{L}(W^z)$ has a density ρ with respect to the Lebesgue measure. Let us first discuss the connection between our general framework and ρ.

Lemma 5.1. *Let (W, W', G) be a Stein coupling. Then, with $\hat{K}(t)$ as in (2.6),*

$$\mathbb{E}\{G(\mathrm{I}[W \leqslant u < W'] - \mathrm{I}[W' \leqslant u < W])\} = \mathbb{E}\hat{K}(u - W) = \rho(u) \qquad (5.2) \quad \{73\}$$

for Lebesgue almost all $u \in \mathbb{R}$.

Proof. The first equality is clear. To prove the second one, let f be a bounded Lipschitz-continuous function. We have

$$Gf(W') - Gf(W) = \int_{\mathbb{R}} f'(W + t)\hat{K}(t)dt = \int_{\mathbb{R}} f'(u)\hat{K}(u - W)du,$$

so that, from (1.8) and using Fubini's Theorem,

$$\mathbb{E}\{Wf(W)\} = \int_{\mathbb{R}} f'(u)\mathbb{E}\hat{K}(u - W)du. \qquad (5.3) \quad \{74\}$$

As we may take $f(x) = \delta^{-1} \int_0^\delta \mathrm{I}[x \leqslant a] \leqslant 1$ and thus $f'(x) = \delta^{-1}I[a \leqslant x \leqslant a + \delta]$ for any $a \in \mathbb{R}$ and $\delta > 0$, we have from (5.1) and (5.3) that

$$\int_a^{a+\delta} \mathbb{E}\hat{K}(u - W)du = \int_a^{a+\delta} \rho(u)du$$

which proves the claim as a and δ are arbitrary. $\qquad \square$

In Goldstein and Reinert (1997) and Goldstein and Reinert (2005b) a method was introduced to construct $\mathscr{L}(W^z)$ using an underlying exchangeable pair satisfying the linearity condition (3.2) with $R = 0$. Although the construction itself does not directly lead to a coupling of W^z with W (which is what we ultimately want), it can nevertheless suggest ways to find such couplings; see for example Goldstein (2005) and Ghosh (2009).

We can generalize the idea to our setting. However, as some of the involved measures may become signed measures, we have to proceed with more care. Denote by F the probability measure on $\mathbb{R}^2$ induced by (W, W'). With

$$\varphi(w, w') := \mathbb{E}(GD \mid W = w, W = w') = (w' - w)\mathbb{E}(G \mid W = w, W = w')$$

define a new (possibly signed) measure

$$d\hat{F}(w, w') = \varphi(w, w')dF(w, w').$$

We have $\int_{\mathbb{R}^2} d\hat{F}(w, w') = 1$ because $\mathbb{E}(GD) = 1$ from (1.8). Let now the space $\hat{\Omega} := \mathbb{R}^2 \times [0, 1]$ be equipped with the standard Borel σ-algebra and define the measure $Q = \hat{F} \otimes \ell$ where ℓ is the Lebesgue measure on $[0, 1]$. Define the mapping $W^z : \hat{\Omega} \to \mathbb{R}$ as

$$W^z(w, w', u) := uw' + (1 - u)w$$

Clearly, W^z is measurable. We now have the following.

Lemma 5.2. *Assume that* (1.8) *holds. Then, the measure P^z on $\mathbb{R}$ induced by the mapping W^z is a probability measure and for every function $f : \mathbb{R} \to \mathbb{R}$ with bounded derivative we have*

$$\mathbb{E}\{Wf(W)\} = \int_{\hat{\Omega}} f'(W^z(w,w',u))dQ(w,w',u) = \int_{\mathbb{R}} f'(x)\rho(x)dx. \quad (5.4) \quad \{75\}$$

Proof. Let us proof the first equality of (5.4).

$$\int_{\hat{\Omega}} f'(W^z(w,w',u))dQ(w,w',u)$$

$$= \int_{\mathbb{R}^2} \int_{[0,1]} f'(w + u(w' - w))du\, d\hat{F}(w,w')$$

$$= \int_{\mathbb{R}^2} \frac{f(w') - f(w)}{w' - w} d\hat{F}(w,w')$$

$$= \int_{\mathbb{R}^2} \varphi(w,w') \frac{f(w') - f(w)}{w' - w} dF(w,w')$$

$$= \mathbb{E}\{G(f(W') - f(W))\} = \mathbb{E}\{Wf(W)\}.$$

We thus have proved that the measure P^z, induced by W^z and Q, satisfies

$$\mathbb{E}\{Wf(W)\} = \int_{\mathbb{R}} f'(x)dP^z(x). \quad (5.5) \quad \{76\}$$

Using the special functions from the proof of Lemma 5.1, it is clear from (5.5) and (5.1) that ρ is the Radon-Nykodim derivative of P^z with respect to the Lebesgue measure. This implies the second equality in (5.4). $\qquad \square$

In the case where $\varphi(W,W') \geqslant 0$ almost surely, $\hat{F}$ is also a probability measure and we can write

$$W^z = U\hat{W}' + (1 - U)\hat{W}$$

where (W,W') has distribution $\hat{F}$ and $U \sim U[0,1]$ is independent of (W,W').

6. Proofs of main results

6.1. **Preliminaries.** For a random variable X define the truncated version $\bar{X} := (X \wedge 1) \vee (-1)$. For a real number t define $t_+ = t \vee 0$ and $t_- = t \wedge 0$.

Stein's method for normal approximation is based on the differential equation

$$f'(w) - wf(w) = h(w) - \mathbb{E}h(Z) \quad (6.1) \quad \{77\}$$

where $Z \sim N(0,1)$ which can be solved for any measurable function h for which $\mathbb{E}h(Z)$ exists. The solution f_h is differentiable and, if h is Lipschitz, also f_h' is Lipschitz; see Stein (1986).

Lemma 6.1. *Let W, W', W'', $\tilde{D}$ and G be square integrable random variables. Let h be a measurable function for which $\mathbb{E}h(W)$ and $\mathbb{E}h(Z)$ exist and let f be the solution to* (6.1). *Then*

$$\big|\mathbb{E}h(W) - \mathbb{E}h(Z)\big| \leqslant (\|f\| \vee \|f'\|)r_0 + \|f'\|(r_1 + r_2 + r_3) + |\mathbb{E}R_1(f)| + |\mathbb{E}R_2(f)|,$$

where

$$R_1(f) = (G\tilde{D} - S)(f'(W'') - f'(W)) \quad (6.2) \quad \{78\}$$

$$R_2(f) = G \int_0^D \big(f'(W+t) - f'(W)\big). \qquad (6.3) \quad \{79\}$$

Proof. Let $f = f_h$ be the solution to (6.1). We can assume that $\|f\|$ and $\|f'\|$ are finite otherwise the statement is trivial. From the fundamental theorem of calculus, we have

$$f(W') - f(W) = \int_0^D f'(W+t)dt. \qquad (6.4) \quad \{80\}$$

Multiplying (6.4) by G and comparing it with the left hand side of (6.1) we have

$$
\begin{aligned}
h(W) - \mathbb{E}h(W) &= f'(W) - Wf(W) \\
&= Gf(W') - Gf(W) - Wf(W) \\
&\quad + (S - G\tilde{D})f'(W'') \\
&\quad + (1 - S)f'(W) \\
&\quad + G(\tilde{D} - D)f'(W) \\
&\quad + (S - G\tilde{D})(f'(W) - f'(W'')) \\
&\quad - G \int_0^D \big(f'(W+t) - f'(W)\big)dt.
\end{aligned}
\qquad (6.5) \quad \{82\}
$$

Taking expectation, the lemma is immediate. $\qquad \square$

6.2. **Bound on the Wasserstein distance.** If h is Lipschitz continuous, then the solution f to (6.1) is differentiable, f' is Lipschitz and we have the following bounds:

$$\|f\| \leqslant 2\|h'\|, \quad \|f'\| \leqslant \sqrt{\frac{2}{\pi}}\|h'\|, \quad \|f''\| \leqslant 2\|h'\|. \qquad (6.6) \quad \{83\}$$

(see Stein (1972) and Raič (2004)).

Proof of Theorem 2.1. The following bounds are easy to obtain using Taylor's theorem:

$$|\mathbb{E}R_1(f)| \leqslant 2r_4'\|f'\| + r_5'\|f''\|, \qquad |\mathbb{E}R_2(f)| \leqslant 2r_4\|f'\| + 0.5r_5\|f''\|.$$

Combining this with Lemma 6.1 and the bounds (6.6) for Lipschitz h with $\|h'\| \leqslant 1$ proves the theorem. $\qquad \square$

6.3. **Bounds on the Kolmogorov distance.** If, for some $\varepsilon > 0$, h is of the form

$$
h_{a,\varepsilon}(x) = \begin{cases}
1, & \text{if } x \leqslant a, \\
1 + (a - x)/\varepsilon, & \text{if } a \leqslant x \leqslant a + \varepsilon, \\
0, & \text{if } x > a + \varepsilon,
\end{cases}
\qquad (6.7) \quad \{84\}
$$

then, from pages 23 and 24 of Stein (1986) (see also Chen and Shao (2004)) we have for $f_{a,\varepsilon}$ for every $w, v \in \mathbb{R}$ the bounds

$$0 \leqslant f_{a,\varepsilon}(w) \leqslant \sqrt{2\pi}/4, \qquad |f'_{a,\varepsilon}(w)| \leqslant 1, \qquad |f'_{a,\varepsilon}(w) - f'_{a,\varepsilon}(v)| \leqslant 1 \qquad (6.8) \quad \{85\}$$

and, in addition for every $s, t \in \mathbb{R}$,

$$|f'_{a,\varepsilon}(w+s) - f'_{a,\varepsilon}(w+t)|$$

$$\leqslant (|w| + 1)\min(|s| + |t|, 1) + \varepsilon^{-1} \int_{s \wedge t}^{s \vee t} \mathrm{I}[a \leqslant w + u \leqslant a + \varepsilon] du \quad (6.9) \quad \{86\}$$

$$\leqslant (|w| + 1)\min(|s| + |t|, 1) + \mathrm{I}[a - s \vee t \leqslant w \leqslant a - s \wedge t + \varepsilon]. \quad (6.10) \quad \{87\}$$

Throughout this section let $\kappa := d_{\mathrm{K}}\big(\mathscr{L}(W), \mathrm{N}(0,1)\big)$. Now, it is not difficult to see that for any $\varepsilon > 0$,

$$\kappa \leqslant \sup_{a \in \mathbb{R}} \big| \mathbb{E}h_{a,\varepsilon}(W) - \mathbb{E}h_{a,\varepsilon}(Z) \big| + 0.4\varepsilon, \qquad (6.11) \quad \{90\}$$

so that we can use Lemma 6.1 for functions of the form (6.7). If $f_{a,\varepsilon}$ is the solution to (6.1), from the bounds (6.8) we thus have

$$\kappa \leqslant r_0 + r_1 + r_2 + r_3 + 0.4\varepsilon + \sup_{a \in \mathbb{R}} |\mathbb{E}R_1(f_{a,\varepsilon})| + \sup_{a \in \mathbb{R}} |\mathbb{E}R_2(f_{a,\varepsilon})|, \quad (6.12) \quad \{91\}$$

which will be the basis for further bounds on κ.

Recall the well known relation between the arithmetic and geometric mean, that is, for each $x, y > 0$ and $\theta > 0$ we have

$$\sqrt{xy} \leqslant \frac{\theta x + \theta^{-1} y}{2}, \qquad (6.13) \quad \{92\}$$

which will be used several times. We will also make use of the following simple lemma in order to implement the recursive approach; see Raič (2003).

Lemma 6.2. *For any random variable V and for any $a < b$ we have*

$$\mathbb{P}[a \leqslant V \leqslant b] \leqslant \frac{b - a}{\sqrt{2\pi}} + 2\, d_{\mathrm{K}}\big(\mathscr{L}(V), \mathrm{N}(0,1)\big). \qquad (6.14) \quad \{93\}$$

Proof of Theorem 2.5. Let $f = f_{a,\varepsilon}$ be the solution to (6.1), where we will from now on omit the dependency on a and ε for better readability. Let $I_1 = \mathrm{I}[|G| \leqslant \alpha, |\tilde{D}| \leqslant \tilde{\beta}, |D'| \leqslant \beta', |S| \leqslant \gamma]$ and write (6.2) as

$$\mathbb{E}R_1(f) = \mathbb{E}\big\{ (G\tilde{D} - S)(1 - I_1)(f'(W'') - f'(W)) \big\}$$
$$+ \mathbb{E}\big\{ (G\tilde{D} - S)I_1(f'(W'') - f'(W)) \big\} =: J_1 + J_2.$$

Using (6.8), the bound $|J_1| \leqslant r_6'$ is immediate. Let for convenience $k := \mathbb{E}|W| + 1$. Using (6.9) and Lemma 6.2,

$$J_2 \leqslant \mathbb{E}\big| (G\tilde{D} - S)I_1\big(f'(W'') - f'(W)\big)\big|$$

$$\leqslant (\alpha\tilde{\beta} + \gamma)k\beta' + (\alpha\tilde{\beta} + \gamma)\varepsilon^{-1} \int_{-\beta'}^{\beta'} \mathbb{P}[a \leqslant W + u \leqslant a + \varepsilon] du$$

$$\leqslant (\alpha\tilde{\beta} + \gamma)k\beta' + 0.8(\alpha\tilde{\beta} + \gamma)\beta' + 4(\alpha\tilde{\beta} + \gamma)\beta'\varepsilon^{-1}\kappa.$$

Similarly, let $I_2 = \mathrm{I}[|G| \leqslant \alpha, |D| \leqslant \beta]$ and write (6.3) as

$$\mathbb{E}R_2(f) = \mathbb{E}\Big\{ G(1 - I_2) \int_0^D \big(f'(W + t) - f'(W)\big) dt \Big\}$$

$$+ \mathbb{E}\Big\{ GI_2 \int_0^D \big(f'(W + t) - f'(W)\big) dt \Big\} =: J_3 + J_4.$$

By (6.8), $|J_3| \leqslant r_6$. Using (6.9) and Lemma 6.2,

$$J_4 \leqslant \mathbb{E}\Big| GI_2 \int_{D_-}^{D_+} \big|f'(W+t) - f'(W)\big| dt \Big|$$

$$\leqslant \alpha \int_{-\beta}^{\beta} |t| k dt + \alpha \varepsilon^{-1} \int_{-\beta}^{\beta} \int_{t_-}^{t_+} \mathbb{P}[a \leqslant W + u \leqslant a + \varepsilon] du dt$$

$$\leqslant \alpha\beta^2 k + 0.4\alpha\beta^2 + 2\alpha\beta^2\varepsilon^{-1}\kappa,$$

so that, collecting all the bounds, setting $\varepsilon = 4\alpha\beta^2 + 8(\alpha\tilde{\beta}+\gamma)\beta'$ and making use of (6.12), we obtain

$$\kappa \leqslant r_0 + r_1 + r_2 + r_3 + r_6 + r_6' + (\alpha\tilde{\beta} + \gamma)(k + 0.8)\beta' + (k + 0.4)\alpha\beta^2$$

$$+ 0.4\varepsilon + \big(2\alpha\beta^2 + 4(\alpha\tilde{\beta} + \gamma)\beta'\big)\varepsilon^{-1}\kappa$$

$$= r_0 + r_1 + r_2 + r_3 + r_6 + r_6' + (\alpha\tilde{\beta} + \gamma)(k + 4)\beta' + (k + 2)\alpha\beta^2 + 0.5\kappa$$

which, solving for κ, proves the theorem. □

Proof of Theorem 2.8. Let $f = f_{a,\varepsilon}$ be the solution to (6.1). As we assume $W'' = W$, we have $R_1(f) = 0$. Write (6.3) as

$$R_2(f) = \mathbb{E} \int_{-\infty}^{\infty} \big(f'(W+t) - f'(W)\big) K^W(t) dt$$

$$= \mathbb{E} \int_{|t|>1} \big(f'(W+t) - f'(W)\big) K^W(t) dt$$

$$+ \mathbb{E} \int_{|t|\leqslant 1} \big(f'(W+t) - f'(W)\big)\big(K^W(t) - K(t)\big) dt$$

$$+ \mathbb{E} \int_{|t|\leqslant 1} \big(f'(W+t) - f'(W)\big) K(t) dt$$

$$=: J_1 + J_2 + J_3.$$

Clearly, by (6.8), $|J_1| \leqslant r_4$. Now let us bound J_2. By (6.10),

$$|J_2| \leqslant \mathbb{E} \int_{|t|\leqslant 1} \big(|W| + 1\big)|t||K^W(t) - K(t)| dt$$

$$+ \mathbb{E} \int_0^1 \mathrm{I}[a - t \leqslant W \leqslant a + \varepsilon]|K^W(t) - K(t)| dt$$

$$+ \mathbb{E} \int_{-1}^0 \mathrm{I}[a \leqslant W \leqslant a - t + \varepsilon]|K^W(t) - K(t)| dt$$

$$=: J_{2,1} + J_{2,2} + J_{2,3},$$

Using (6.13), we have for any $\theta > 0$ that

$$J_{2,1} \leqslant \frac{\theta}{2}\mathbb{E} \int_{|t|\leqslant 1} (|W| + 1)^2 |t| dt + \frac{1}{2\theta}\mathbb{E} \int_{|t|\leqslant 1} |t||K^W(t) - K(t)|^2 dt$$

$$= \frac{\theta}{2}\mathbb{E}(|W| + 1)^2 + \frac{1}{2\theta}r_8^2.$$

Let $\alpha_0 = (\mathbb{E}(|W| + 1)^2)^{1/2}$ and choose $\theta = \alpha_0^{-1}r_8$, so that

$$J_{2,1} \leqslant \alpha_0 r_8.$$

Let now $\delta = 0.4\varepsilon + 2\kappa$. Then, from Lemma 6.2, we have for $t > 0$

$$\mathbb{P}[a - t \leqslant W \leqslant a + \varepsilon] \leqslant \delta + 0.4t. \qquad (6.15) \quad \{94\}$$

Using (6.13),

$$J_{2,2} \leqslant \mathbb{E}\left\{\int_0^1 \left[0.5\varepsilon\left(\delta + 0.4t\right)^{-1}\mathrm{I}[a - t \leqslant W \leqslant a + \varepsilon]\right.\right.$$
$$\left.\left. + 0.5\varepsilon^{-1}\left(\delta + 0.4t\right)\left(K^W(t) - K(t)\right)^2\right] dt\right\}$$
$$\leqslant 0.5\varepsilon + 0.5\varepsilon^{-1}\delta \int_0^1 \mathrm{Var}(K^W(t))dt + 0.2\varepsilon^{-1}\int_0^1 t\,\mathrm{Var}(K^W(t))dt.$$

A similar bound holds for $J_{2,3}$ so that

$$|J_2| \leqslant \alpha_0 r_8 + \varepsilon + 0.5\varepsilon^{-1}\delta r_7 + 0.2\varepsilon^{-1}r_8^2.$$

By (6.9),

$$|J_3| \leqslant \mathbb{E}\int_{|t|\leqslant 1} (|W| + 1)|tK(t)|dt$$
$$+ \varepsilon^{-1}\int_{|t|\leqslant 1}\left|\int_0^t \mathbb{P}[a \leqslant W + u \leqslant a + \varepsilon]du\right||K(t)|dt$$
$$\leqslant (\mathbb{E}|W| + 1)r_5 + \varepsilon^{-1}\int_{|t|\leqslant 1} \delta|tK(t)|dt \leqslant (\mathbb{E}|W| + 1)r_5 + \varepsilon^{-1}\delta r_5.$$

Choose now

$$\varepsilon = \sqrt{1.4}\big[\kappa(2r_5 + r_7) + 0.2r_8^2\big]^{1/2}.$$

From (6.12) (note that $r_1 = 0$ because $\tilde{D} = D$), using that $\sqrt{x+y} \leqslant \sqrt{x} + \sqrt{y}$ and (6.13), we obtain

$$\kappa \leqslant r_0 + r_2 + r_3 + 0.4\varepsilon + |J_1| + |J_2| + |J_3|$$
$$\leqslant r_0 + r_2 + r_3 + r_4 + (\mathbb{E}|W| + 1)r_5 + \alpha_0 r_8 + 1.4\varepsilon$$
$$\quad + \varepsilon^{-1}(\delta(r_5 + 0.5r_7) + 0.2r_8^2)$$
$$\leqslant r_0 + r_2 + r_3 + r_4 + (\mathbb{E}|W| + 1.4)r_5 + 0.2r_7 + \alpha_0 r_8 + 1.4\varepsilon$$
$$\quad + \varepsilon^{-1}\big(\kappa(2r_5 + r_7) + 0.2r_8^2\big)$$
$$\leqslant r_0 + r_2 + r_3 + r_4 + (\mathbb{E}|W| + 1.4)r_5 + 0.2r_7 + \alpha_0 r_8$$
$$\quad + 2.4\big(\kappa(2r_5 + r_7) + 0.2r_8^2\big)^{1/2}$$
$$\leqslant r_0 + r_2 + r_3 + r_4 + (\mathbb{E}|W| + 1.4)r_5 + 0.2r_7 + (\alpha_0 + 1.1)r_8$$
$$\quad + 2.4\big(\kappa(2r_5 + r_7)\big)^{1/2}$$
$$\leqslant r_0 + r_2 + r_3 + r_4 + (\mathbb{E}|W| + 1.4 + 2.4\theta^{-1})r_5 + (0.2 + 1.2\theta^{-1})r_7$$
$$\quad + (\alpha_0 + 1.1)r_8 + 1.2\theta\kappa.$$

Choosing $\theta = 1/2.4$ and solving for κ proves the claim. $\qquad\square$

Proof of Theorem 2.9. Let $f = f_{a,\varepsilon}$ be the solution to (6.1). Now, from (6.9),

$$|R_1(f)| \leqslant \left|(S - G\tilde{D})(|W| + 1)\bar{D}'\right|$$
$$+ |S - G\tilde{D}|\varepsilon^{-1}\int_{D'_-}^{D'_+} \mathrm{I}[a \leqslant W + u \leqslant a + \varepsilon]du$$

so that

$$\mathbb{E}|R_1(f)| \leqslant \mathbb{E}\left|(S - G\tilde{D})(|W| + 1)\bar{D}'\right|$$
$$+ \varepsilon^{-1}\mathbb{E}\left|(S - G\tilde{D})D'S_\varepsilon(\mathscr{L}(W|G, \tilde{D}, D'))\right|.$$

Furthermore,

$$|R_2(f)| \leqslant |G|\int_{D_-}^{D_+}(|W| + 1)(|t| \wedge 1)dt$$
$$+ \varepsilon^{-1}|G|\int_{D_-}^{D_+}\int_{t_-}^{t_+} \mathrm{I}[a \leqslant W + u \leqslant a + \varepsilon]dudt$$

hence

$$\mathbb{E}|R_2(f)| \leqslant 0.5\mathbb{E}|G(|W| + 1)\bar{D}^2| + 0.5\varepsilon^{-1}\mathbb{E}\left|GD^2S_\varepsilon(\mathscr{L}(W|G, D))\right|$$

Combining these bounds with (6.12) proves the theorem. $\qquad\square$

To prove Lemma 2.10 we first need a simple lemma.

Lemma 6.3. *Let $\beta_{k,l}$, $1 \leqslant l < k$, $k = 2, \ldots, n$, be non-negative real numbers. If $a_1 \leqslant b_1 = 1$ and if there are constants $q \geqslant 1$ and $p < 1$ such that, for all $k = 2, 3, \ldots, n$, we have $\sum_{l=1}^{k-1}\beta_{k,l} \leqslant p$,*

$$a_k = q + \sum_{l=1}^{k-1}\beta_{k,l}a_l \qquad and \qquad b_k = q + pb_{k-1},$$

then $a_k \leqslant b_k \leqslant q/(1-p)$ for all $1 \leqslant k \leqslant n$.

Proof. Note first that $q/(1-p) \geqslant 1$ and

$$b_k = \left(1 - \frac{q}{1-p}\right)p^k + \frac{q}{1-p},$$

hence $b_1, b_2, \ldots$ is increasing with upper bound $q/(1-p)$. The proof of $a_k \leqslant b_k$ is now a simple induction on k. By assumption $a_1 \leqslant b_1$, which verifies the base case. Using that $a_l \leqslant b_l$ for all $1 \leqslant l \leqslant k$, we have

$$a_{k+1} = q + \sum_{l=1}^{k}\beta_{k+1,l}a_l \leqslant q + \sum_{l=1}^{k}\beta_{k+1,l}b_l$$
$$\leqslant q + \sum_{l=1}^{k}\beta_{k+1,l}b_k \leqslant q + pb_k = b_{k+1}. \qquad\square$$

Proof of Lemma 2.10. First note that (2.8) still holds if we replace A by $\bar{A} := A \vee 1$. Let $\sigma_1 := 1$ and define a new sequence $a_k = \sigma_k\kappa_k$ for $1 \leqslant k \leqslant n$.

Using inequality (2.8) with $\varepsilon = c_n \alpha_n / \sigma_k$ for a constant $c_n > 1$ to be chosen later,

$$a_k \leqslant \bar{A} + 0.4 c_n \alpha_n + \sum_{l=1}^{k-1} \beta_{k,l} a_l$$

for all $1 \leqslant k \leqslant n$, where

$$\beta_{k,l} = \frac{\sigma_k A_{k,l}}{c_n \sigma_l \alpha_n}.$$

Note that $\sum_{l=1}^{k-1} \beta_{k,l} \leqslant c_n^{-1}$. Consider now the solution b_k to the recursive equation

$$b_1 = 1; \qquad b_k = \bar{A} + 0.4 c_n \alpha_n + c_n^{-1} b_{k-1}, \quad \text{for } 2 \leqslant k \leqslant n.$$

Note that $\bar{A} + 0.4 c_n \alpha_n \geqslant 1$, hence we obtain from Lemma 6.3 that

$$a_n \leqslant b_n \leqslant \frac{c_n(\bar{A} + 0.4 c_n \alpha_n)}{c_n - 1}.$$

Minimizing over $c_n > 1$ we can chose

$$c_n = 1 + \frac{\sqrt{2\alpha_n(2\alpha_n + 5\bar{A}))}}{2\alpha_n} = 1 + \frac{\alpha_n'}{2\alpha_n}.$$

Recalling that $\kappa_n = a_n / \sigma_n$, the claim follows. $\qquad\square$

ACKNOWLEDGMENTS

REFERENCES

P. Baldi, Y. Rinott, and C. Stein (1989). A normal approximation for the number of local maxima of a random function on a graph. In *Probability, statistics, and mathematics*, pages 59–81. Academic Press, Boston, MA.

A. D. Barbour (1988). Stein's method and Poisson process convergence. *J. Appl. Probab.* **25A**, 175–184.

A. D. Barbour (2009). Univariate approximations in the infinite occupancy scheme. *ALEA* **6**, 415—433.

A. D. Barbour and G. K. Eagleson (1986). Random association of symmetric arrays. *Stochastic Anal. Appl.* **4**, 239–281.

A. D. Barbour, M. Karoński, and A. Ruciński (1989). A central limit theorem for decomposable random variables with applications to random graphs. *J. Combin. Theory Ser. B* **47**, 125–145.

A. D. Barbour, L. Holst, and S. Janson (1992). *Poisson approximation*, volume 2 of *Oxford Studies in Probability*. The Clarendon Press Oxford University Press, New York. ISBN 0-19-852235-5. Oxford Science Publications.

H. Bergström (1944). On the central limit theorem. *Skand. Aktuarietidskr.* **27**, 139–153.

E. Bolthausen (1982). Exact convergence rates in some martingale central limit theorems. *Ann. Probab.* **10**, 672–688.

E. Bolthausen (1984). An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrsch. Verw. Gebiete* **66**, 379–386.

M. V. Boutsikas and M. V. Koutras (2002). On the number of overflown urns and excess balls in an allocation model with limited urn capacity. *J. Statist. Plann. Inference* **104**, 259–286.

T. Cacoullos, V. Papathanasiou, and S. A. Utev (1994). Variational inequalities with examples and an application to the central limit theorem. *Ann. Probab.* **22**, 1607–1618.

S. Chatterjee (2005). *Concentration inequalities with exchangeable pairs*. PhD thesis, Stanford University.

S. Chatterjee (2007). Stein's method for concentration inequalities. *Probab. Theory Related Fields* **138**, 305–321.

S. Chatterjee (2008). A new method of normal approximation. *Ann. Probab.* **36**, 1584–1610.

S. Chatterjee (2009). Fluctuations of eigenvalues and second order poincar inequalities. *Probab. Theory Related Fields* **143**, 1–40.

S. Chatterjee, J. Fulman, and A. Röllin (2006). Exponential approximation by exchangeable pairs and spectral graph theory. *Preprint.* Available at `http://arxiv.org/abs/math/0605552`.

L. H. Y. Chen (1975). Poisson approximation for dependent trials. *Ann. Probability* **3**, 534–545.

L. H. Y. Chen (1998). Stein's method: some perspectives with applications. In *Probability towards 2000 (New York, 1995)*, volume 128 of *Lecture Notes in Statist.*, pages 97–122. Springer, New York.

L. H. Y. Chen and Q.-M. Shao (2001). A non-uniform Berry-Esseen bound via Stein's method. *Probab. Theory Related Fields* **120**, 236–254.

L. H. Y. Chen and Q.-M. Shao (2004). Normal approximation under local dependence. *Ann. Probab.* **32**, 1985–2028.

L. H. Y. Chen and Q.-M. Shao (2005). Stein's method for normal approximation. In *An introduction to Stein's method*, volume 4 of *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.*, pages 1–59. Singapore Univ. Press, Singapore.

A. Dembo and Y. Rinott (1996). Some examples of normal approximations by Stein's method. In *Random discrete structures (Minneapolis, MN, 1993)*, volume 76 of *IMA Vol. Math. Appl.*, pages 25–44. Springer, New York.

P. Diaconis (1977). The distribution of leading digits and uniform distribution mod 1. *Ann. Probability* **5**, 72–81.

P. Diaconis and S. Zabell (1991). Closed form summation for classical distributions: variations on a theme of de Moivre. *Statist. Sci.* **6**, 284–302.

R. Durrett (2007). *Random graph dynamics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge. ISBN 978-0-521-86656-9; 0-521-86656-1.

G. Englund (1981). A remainder term estimate for the normal approximation in classical occupancy. *Ann. Probab.* **9**, 684–692.

J. Fulman (2004a). Stein's method and non-reversible Markov chains. In *Stein's method: expository lectures and applications*, volume 46 of *IMS Lecture Notes Monogr. Ser.*, pages 69–77. Inst. Math. Statist., Beachwood, OH.

J. Fulman (2004b). Stein's method, Jack measure, and the Metropolis algorithm. *J. Combin. Theory Ser. A* **108**, 275–296.

S. Ghosh (2009). L^p bounds for a combinatorial central limit theorem with involutions. *Preprint.*

S. Ghosh (2010). Multivariate concentration of measure type results using exchangeable pairs and size biasing. *Preprint.* Available at `arXiv:1001.1396v1`.

A. L. Gibbs and F. E. Su (2002). On choosing and bounding probability metrics. *International Statistical Review / Revue Internationale de Statistique* **70**, 419–435.

L. Goldstein (2005). Berry-Esseen bounds for combinatorial central limit theorems and pattern occurrences, using zero and size biasing. *J. Appl. Probab.* **42**, 661–683.

L. Goldstein (2010). A Berry-Esseen bound with applications to counts in the Erdős-Rényi random graph. Available at `http://arxiv.org/abs/1005.4390`.

L. Goldstein and M. D. Penrose (to appear). Normal approximation for coverage models over binomial point processes. *Ann. Appl. Probab.*

L. Goldstein and G. Reinert (1997). Stein's method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.* **7**, 935–952.

L. Goldstein and G. Reinert (2005a). Distributional transformations, orthogonal polynomials, and Stein characterizations. *J. Theoret. Probab.* **18**, 237–260.

L. Goldstein and G. Reinert (2005b). Zero biasing in one and higher dimensions, and applications. In *Stein's method and applications*, volume 5 of *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.*, pages 1–18. Singapore Univ. Press.

L. Goldstein and Y. Rinott (1996). Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Probab.* **33**, 1–17.

F. Götze (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19**, 724–739.

S. T. Ho and L. H. Y. Chen (1978). An L_p bound for the remainder in a combinatorial central limit theorem. *Ann. Probability* **6**, 231–249.

H.-K. Hwang and S. Janson (2008). Local limit theorems for finite and infinite urn models. *Ann. Probab.* **36**, 992–1022.

S. Janson and M. J. Luczak (2008). Susceptibility in subcritical random graphs. *J. Math. Phys.* **49**, 125207, 23.

N. L. Johnson and S. Kotz (1977). *Urn models and their application.* John Wiley & Sons, New York-London-Sydney. An approach to modern discrete probability theory, Wiley Series in Probability and Mathematical Statistics.

V. F. Kolchin, B. A. Sevast′yanov, and V. P. Chistyakov (1978). *Random allocations.* V. H. Winston & Sons, Washington, D.C.

W. Y. Loh (1975). On the normal approximation for sums of mixing random variables. Master's thesis, University of Singapore.

H. M. Luk (1994). *Stein's method for the gamma distribution and related statistical applications.* PhD thesis, University of Southern California.

A. M. Makowski and A. Shwartz (1994). On the Poisson equation for count-
able Markov chains: existence of solutions and parameter dependence by
probabilistic methods. Technical report, Institute for Systems Research.

E. Meckes (2008). Linear functions on the classical matrix groups. *Trans.
Amer. Math. Soc.* **360**, 5355–5366.

I. Nourdin and G. Peccati (2009). Stein's method on wiener chaos. *Probab.
Theory Related Fields.*

I. Nourdin, G. Peccati, and G. Reinert (2009). Second order Poincaré in-
equalities and CLTs on Wiener space. *J. of Funct. Anal.* **257**, 593–609.

M. D. Penrose and J. E. Yukich (2001). Central limit theorems for some
graphs in computational geometry. *Ann. Appl. Probab.* **11**, 1005–1041.

M. Raič (2003). Normal approximation with Stein's method. In *Proceedings
of the Seventh Young Statisticians Meeting.*

M. Raič (2004). A multivariate CLT for decomposable random vectors with
finite second moments. *J. Theoret. Probab.* **17**, 573–603.

G. Reinert (1998). Couplings for normal approximations with Stein's
method. In *Microsurveys in discrete probability (Princeton, NJ, 1997)*,
volume 41 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages
193–207. Amer. Math. Soc., Providence, RI.

G. Reinert and A. Röllin (2009a). Multivariate normal approximation with
Stein's method of exchangeable pairs under a general linearity condition.
Ann. Probab. **37**, 2150–2173.

G. Reinert and A. Röllin (2009b). *U*-statistics and random subgraph counts:
Multivariate normal approximation via exchangeable pairs and embed-
ding. *Preprint.* Available at `arXiv:0912.3425v1`.

I. Rinott and V. I. Rotar′ (1998). Some estimates for the rate of convergence
in the CLT for martingales. I. *Theory Probab. Appl.* **43**, 604–619.

Y. Rinott and V. Rotar′ (1997). On coupling constructions and rates in the
CLT for dependent summands with applications to the antivoter model
and weighted *U*-statistics. *Ann. Appl. Probab.* **7**, 1080–1105.

Y. Rinott and V. Rotar′ (2000). Normal approximations by Stein's method.
Decis. Econ. Finance **23**, 15–29.

A. Röllin (2007a). On Stein factors and the construction of examples with
sharp rates in Stein's method. *Preprint.*

A. Röllin (2007b). Translated Poisson approximation using exchangeable
pair couplings. *Ann. Appl. Probab.* **17**, 1596–1614.

A. Röllin (2008a). A note on the exchangeability condition in Stein's method.
Statist. Probab. Lett. **78**, 1800–1806.

A. Röllin (2008b). Symmetric binomial approximation for sums of locally
dependent random variables. *Electron. J. Probab.* **13**, 756–776.

N. Ross (to appear). Step size in Stein's method of exchangeable pairs.
Combinatorics, Probability, and Computing.

Q.-M. Shao (2005). An explicit Berry-Esseen bound for Student's *t*-statistic
via Stein's method. **5**, 143–155.

Q.-M. Shao and Z.-G. Su (2006). The Berry-Esseen bound for character
ratios. *Proc. Amer. Math. Soc.* **134**, 2153–2159 (electronic).

C. Stein (1972). A bound for the error in the normal approximation to the
distribution of a sum of dependent random variables. In *Proceedings of*

the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory, pages 583–602, Berkeley, Calif. Univ. California Press.

C. Stein (1986). *Approximate computation of expectations.* Institute of Mathematical Statistics Lecture Notes—Monograph Series, 7. Institute of Mathematical Statistics, Hayward, CA.

C. Stein (1995). The accuracy of the normal approximation to the distribution of the traces of powers of random orthogonal matrices. Technical Report No. 470, Standford University Department of Statistics.

J. Sunklodas (2008). On L_1 bounds for asymptotic normality of some weakly dependent random variables. *Acta Appl. Math.* **102**, 87–98.

L. Zhao, Z. Bai, C.-C. Chao, and W.-Q. Liang (1997). Error bound in a central limit theorem of double-indexed permutation statistics. *Ann. Statist.* **25**, 2210–2227.